Shaping the Profession:

Generative AI and professional judgement in accounting

Case-evidence from a mid-tier accounting practice

The Research Team

Dr Sung Hwan Chai

Lecturer in Accounting at Alliance Manchester Business School (AMBS)

Principal Investigator

Prof. Brian Nicholson

Professor of Business Information Systems at AMBS

Co-Investigator

Dr George Salijeni

Lecturer in Accounting at Aston Business School

Co-Investigator

Dr Leonid Sokolovskyy

Lecturer in Accounting and Finance at AMBS

Research Associate

Dr Yasmine Chahed

Visiting Research Fellow at AMBS

Research Consultant

The Beever and Struthers Team

Caroline Monk

FCA Chartered Accountant

Executive Partner

John Toon

ACA Chartered Accountant

Technology Lead

Bec Hesketh

ACA Chartered Accountant

Assistant Manager

Beever and Struthers

Beever and Struthers (B&S) is a firm of chartered accountants, business advisors, and leading audit providers. Headquartered in Manchester, the firm is ranked among the Top 10 firms in the Northwest and 53rd in the Accountancy Age Top 50+50 Firms 2024. With over 100 years of experience, 18 expert partners, and a national client base, B&S delivers a comprehensive suite of services (audit, accountancy, tax, corporate finance, internal audit, and advisory) to businesses, not-for-profits, housing providers, education bodies and professional practices across the UK. On 4 August 2025, B&S announced its merger with Menzies LLP, forming a national firm with a combined fee income of £110 million and a workforce exceeding 1,000 professionals. The merger was complete on 1 October 2025.

The firm has received support for its ongoing digital transformation strategy through a strategic Knowledge Transfer Partnership with Aston and Manchester Universities.

Innovate UK Management Knowledge Transfer Partnership (mKTP)

The Digital Audit KTP[1] is a partnership between Alliance Manchester Business School at the University of Manchester, Aston University Business School, and B&S, set up to deliver digital transformation through business process re-engineering of current audit services covering people, processes and the underlying systems and technology.

The Institute of Chartered Accountants of Scotland

The Institute of Chartered Accountants of Scotland (ICAS) is a global professional membership organisation and business network for Chartered Accountants (CAs), established by Royal Charter in 1854. ICAS acts as an educator, examiner, regulator and professional awarding body, supporting the profession through qualification, technical guidance, ethics and public-interest thought leadership. With a worldwide community of over 24,000 members, ICAS members work across public practice and industry in the UK and internationally. ICAS awarded the research team a research grant to specifically look at how GenAI impacts the professional judgement process in an accountancy setting.

[1] The Digital Audit KTP: Transforming Audits with Digital Technology. https://www.digital-audit.uk/

Contents

Introduction

Generative Artificial Intelligence (Gen AI) marks a major technological innovation that is expected to fundamentally redefine the role of accounting professionals.[2] Gen AI includes large language models (LLMs)[3] that are pre-trained on vast volumes of textual data.[4] This allows the generation of sophisticated written outputs, complex analysis, and interactive, natural language exchanges with users. Although the outputs created by LLMs remain rooted in statistical prediction, they depart significantly from the rigid and task-specific AI models of the past, which focused primarily on retrieving and sorting data constrained by predefined rules and classifications.[5] However, while Gen AI technologies are expected to evolve and spread rapidly, it is uncertain how these technologies will impact the professional judgement process of accounting professionals.

The ability to exercise professional judgement is widely described as being central to the role of the accounting professional in delivering value to individuals, businesses, and the wider public.[6] When exercising professional judgement, accountants combine their professional knowledge, skill, and ethics to reach nuanced and informed opinions and decisions. This process is situational and often implied. It sits between technical rules and context-specific requirements and is shaped by standards, evidence requirements, consultation, and documentation practices. Further, professional judgement in accounting is required beyond accounting treatments across a wide range of matters, including ethics, financial and wider corporate reporting, and business decision-making in general. It therefore plays a key part in the day-to-day activities undertaken by an accounting professional, regardless of their role and the sector in which they operate.

Drawing on a year-long action research collaboration with a mid-tier accounting firm that provides a wide range of accounting, audit, and tax services (Beever and Struthers, now part of Menzies LLP), this report investigates, within the context of accounting practices, two questions that are central to understanding the practical opportunities and challenges of Gen AI in shaping the profession:

1.
How does Gen AI impact the professional judgement process of accounting professionals?

2.
What are the ethical implications of using Gen AI when developing and making professional judgements in accounting settings?

The analysis is based on firm-wide surveys, training interventions, fifty semi-structured interviews, participant observation, and four in-depth case studies with volunteer Gen AI innovators that offer insight into how accounting professionals have applied Gen AI. While the findings apply to accounting practices, they also hold relevance for professionals outside of accounting. Each case study highlights distinctive aspects of the human/Gen AI interaction across different types of accounting work in auditing, financial reporting, corporate reporting, and corporate finance advisory.

2 Institute of Chartered Accountants of Scotland (ICAS) (2025). Society First: Shaping the Profession. Available: https://www.icas.com/news-insights-events/documents/society-first-shaping-the-profession-2025
3 Examples include ChatGPT developed by OpenAI, Gemini developed by Google, and Claude developed by Anthropic.
4 GPT stands for Generative Pre-trained Transformer. For images and videos text-to-image models like DALL-E and Midjourney are used, but the principle of training on large datasets remains.
5 LLMs and Gen AI in Audit — The Digital Audit KTP. Available at: https://www.digital-audit.uk/blog/llms-and-genai-in-audit
6 ICAS (2025). Society First: Shaping the Profession. Available: https://www.icas.com/news-insights-events/documents/society-first-shaping-the-profession-2025

Through these practical examples, we illustrate how Gen AI tools relate to the professional judgement process and create both opportunities and risks for accounting professionals in distinctive ways.

Firstly, as a practical takeaway, we show how accounting professionals can assess the potential use of Gen AI by adapting Davenport's (2005) "Classification Structure for Knowledge-intensive Processes"[7] to link Gen AI use to the level of interpretation required and the degree of interdependence involved in completing a work task. This lens highlights that relatively routine and structured tasks that require low levels of recognising patterns in information may be better suited to uses of Gen AI in the context of professional judgement. These could include data extraction, summarisation, or the initial organisation of information. By contrast, tasks that require greater contextual interpretation, judgement under uncertainty, or coordination across people and processes generally require a clearer outline of the Gen AI model's role, stronger controls on outputs, and explicit professional oversight of any output produced by Gen AI before it feeds into the professional's evaluative process.

Secondly, concerning the impact of Gen AI on professional judgement across different use-cases, we show how the experience of Gen AI as an interactive technology (used through iterative prompting and response) creates new pathways. Doing so, Gen AI changes the modes of human-technology interaction, relative to earlier technologies introduced into accounting and auditing, to include supervision, training, and simulated dialogue.

Thirdly, beyond well-recognised issues such as hallucinations (generated outputs that are false, nonsensical, or entirely fabricated, presented as if they were true) and data security, we see examples where Gen AI may shape human judgement indirectly. This can be by framing issues, suggesting lines of reasoning, and anchoring later subsequent evaluations.

For example, as Gen AI adapts to the user's prompts and preferences through repeated use, it increases the risk of cognitive bias in professional judgements. In addition, we observed instances of users attributing human-like features and traits to Gen AI technology ("anthropomorphising" or "humanising"), which increases the risk of 'forgetting' that its outputs remain fundamentally statistical and pattern-based. We also highlight the emerging risk of Gen AI 'feedback loops', in which AI-generated outputs spread into subsequent automated workflows and processes, potentially exacerbating biases, errors, or unexamined assumptions.

Finally, we identify a set of practical benefits and opportunities beyond the automation of routine workflows that practitioners experienced when using Gen AI to support professional judgement, particularly by reducing time spent on knowledge gathering and analysis, as well as the documentation of their judgement. Practitioners were also able to produce concise summaries of relevant accounting standards using Gen AI. We also observed how Gen AI was used as a learning support, which was especially valuable for junior staff and when working in unfamiliar areas.

What becomes clear across all use cases is that the introduction of Gen AI tools to support the professional judgement process requires a strong professional response, including attention to professional ethics, governance, and oversight. In this report, we suggest that professional ethics should be treated as a central pillar for clarifying and reinforcing human agency or autonomy, responsibility, and accountability in Gen AI-supported activity. We argue for preserving time and space for reflection within judgement processes, alongside practical safeguards to ensure Gen AI is used appropriately. In the age of Gen AI, the exercise of professional judgement in accounting remains a human responsibility grounded in professional knowledge, experience, and ethics, and cannot be delegated to computers or algorithms.

[7] Davenport, T. (2005). Thinking for a living: How to get better performance and results from knowledge workers. Harvard Business School Press.

Reinforcing professional judgement in the age of Gen AI

While there is no single, universally accepted definition, professional judgement is often seen as a cognitive and ethical experience and professional standards with personal qualities to form opinions or take decisions"[8] in uncertain contexts. Such situations are common in the work of accounting professionals, where the interpretation and application of generally accepted principles and standards can be flexible and require significant situated awareness, and where professional judgements frequently combine accounting knowledge with managerial, strategic, operational, and governance considerations.[9] Because the act of making a professional judgement is often subjective, it is particularly vulnerable to cognitive biases.[10] The implications of this vulnerability for the quality of and trust in accounting decision-making have driven the demand for professional guidance.

The Institute of Chartered Accountants of Scotland's (ICAS) *Professional Judgement Framework for Financial Reporting Decision Making*[11] identifies four key principles as critical to making a professional accounting judgement in the context of financial reporting: knowledge gathering and analysis, assessment of accounting guidance, following due process in making a judgement, and documentation of the judgement. These principles are supported by a non-exhaustive list of underlying actions accountants might undertake to apply them, such as considering alternatives, considering bias and risks, consulting experts, and engaging in ethical deliberation. Together, they are designed to provide a robust foundation for professional judgement in accounting (Figure 1).

Figure 1: Four pillars of professional judgement (ICAS Framework)[12]

Key principles on which to base professional accounting judgement	**Principle 1** Knowledge gathering and analysis	**Principle 2** Assessment of accounting guidance	**Principle 3** Process for making a judgement	**Principle 4** Documentation of judgement
Assumption: A professional accounting judgement can only be made...	... once all relevant and determinable information has been collected and analysed.	... in the context of the applicable accounting framework, accounting standards, and other relevant literature.	... after undertaking appropriate due process.	... when supported by a clear, contemporaneous record of the information, rationale, process, and any required disclosure or reassessment.

[8] Likierman, A. (2025). Judgement at Work: Making better choices. Profile Books, p. 68.
[9] ICAS (2016). A Professional Judgement Framework for Financial Reporting Decision Making. 2nd edition.
[10] Camilli, R., Cristofaro, M., Hristov, I., & Sargiacomo, M. (2025). Cognitive biases in accounting judgment and decision making: a review, a typology, and a future research agenda. Accounting Forum, 1-30. https://doi.org/10.1080/01559982.2024.2434340
[11] ICAS (2016). A Professional Judgement Framework for Financial Reporting Decision Making. 2nd edition.
[12] Own visualisation based on ICAS (2016). A Professional Judgement Framework for Financial Reporting Decision Making. 2nd edition.

Historically, professional frameworks have reinforced the view that professional judgement is a uniquely human domain. Technology can either support human judgement, for example, by giving us more evidence through big data, or it can make things harder by adding complexity and information overload, which means we rely on human judgement even more.[13]

Gen AI significantly expands the range of tasks that technology can support within the professional judgement process. As it is trained on vast textual datasets and generates responses through probabilistic inference, its capabilities (further enhanced through prompt design[14]) make it well suited to tasks such as reviewing unstructured information, applying relevant standards, refining early evaluations, and documenting actions and reasoning efficiently.

Gen AI models and prompt design

General-purpose Gen AI models:
Models are trained on massive text datasets and produce responses by calculating the most probable output. While they lack human-like understanding, they can perform tasks that imitate analytical reasoning, recognise patterns, and pull information together.

Custom Gen AI models:
Organisations can use a standard Gen AI tool as-is, or a customised version tailored to specific needs, with options such as web search, access to internal data, or organisation-specific instructions enabled or disabled.

Prompt design:
This refers to how users word their requests and provide context so that the Gen AI model produces more relevant, consistent, and useful outputs.

Although it is well established that using technology does not remove the professional's responsibility for the accuracy, integrity, and ethical quality of their work, it is still essential to understand how technology can influence the way professional judgement is formed.[15] This is because the adoption of Gen AI tools by accounting professionals raises important questions. Gen AI will affect the core principles that underpin high-quality and trusted professional judgement. This includes increasing the risk of cognitive bias and creating wider implications for professional expertise, skills, and ethics[16], which we examine throughout the rest of this report.

[13] Salijeni, G., Samsonova-Taddei, A. and Turley, S. (2021). Understanding How Big Data Technologies Reconfigure the Nature and Organization of Financial Statement Audits: A Sociomaterial Analysis. European Accounting Review 30(3): 531-555; Samiolo, R., Spence, C. and Toh, D. (2024). Auditor Judgment in the Fourth Industrial Revolution. Contemporary Accounting Research 41(1): 498-528.
[14] Chen, B., Zhang, Z., Langrené, N., and Zhu, S. (2025). Unleashing the Potential of Prompt Engineering for Large Language Models. Patterns, no. 6 https://doi.org/10.1016/j.patter.2025.101260.
[15] Gunz, S., & Thorne, L. (2020). Thematic symposium: The impact of technology on ethics, professionalism and judgement in accounting. Journal of Business Ethics, 167(2), 153-155.
[16] Tiron-Tudor, A., Rodgers, W. and Deliu, D. (2024). "The Accounting Profession In the twilight Zone: Navigating Digitalisation's Sided Challenges through Ethical Pathways For decision-making." Accounting, Auditing & Accountability Journal 38(3): 990-1018.

Consistent with how the 'profession' is defined within the scope of the ICAS 'Society First: Shaping the Profession' (2025) initiative, we are focused on the judgements made by accountants with a professional background and/or qualification in accounting and who are working in any setting where accounting information and analysis help inform their judgement.

Gen AI and cognitive biases

Because Gen AI can produce rapid, fluent, and confident-sounding answers that can shape what information is considered and how it is evaluated by the professional, the following biases are likely to be especially important.

Automation bias:
The risk of placing too much trust in automated results and failing to verify or challenge them.

Confirmation bias:
The tendency to seek, interpret, and weight information in ways that confirm pre-existing beliefs or hypotheses.

Anchoring bias:
The tendency to rely too heavily on an initial piece of information when making subsequent judgements or estimates.

Unlocking new insights

In investigating the impact of Gen AI on professional judgement, our primary objective was to generate practical insights. To achieve this, we designed a methodology involving two complementary sets of activities, which the research team and the Beever and Struthers (B&S) team delivered in close collaboration between January and October 2025.[17]

Empowering practitioners

The first activity set focused on equipping practitioners to meaningfully incorporate Gen AI into their day-to-day work at B&S, as part of the firm's digital transformation strategy. Led by the B&S team over a 12-week period at the beginning of 2025, activities included a survey sent to all staff to capture existing levels of experience and perceptions of Gen AI use in professional work. The survey achieved a response rate of approximately 70% (out of a population of 292), providing a robust basis for identifying levels of Gen AI use among B&S staff and designing firm-wide training sessions. The initial adoption rate across the firm was as follows: little to no use (40%), some use (cautious and occasional) (26%), regular use (experimenting and experienced) (34%).[18]

While we cannot comment on how representative these results are of practices as a whole, B&S is a mid-tier regional firm that sees itself as being ahead of the curve in adopting new technologies, including Gen AI. It is also important to note that B&S was already positioned as a firm with clear ambitions for technological innovation. The firm has already worked with the academic team for several years through a Knowledge Transfer Partnership (KTP), and its commitment to innovation was underlined by its creation of a stand-alone digital transformation department in 2023.

Following the survey, all staff were invited to participate in optional Gen AI training ("Introduction and basics of Gen AI") as part of the firm's existing digital transformation programme. This was focused on driving the adoption of digital solutions to improve operational quality and efficiency, as well as generating the 'good habits' needed to use Gen AI ethically and responsibly.

The survey also helped identify staff members who were 'innovators' and 'early adopters', who were invited to participate in a pilot phase of the Gen AI programme.[19] They attended focused sessions designed to encourage and support the development of Gen AI use cases. During these sessions, participants were guided through proposing, refining, and experimenting with customising their own Gen AI models (hereafter, custom Gen AI) relevant to their daily work. Participants in groups received feedback prior to finalisation and subsequently presented and evaluated their use cases. In total, eight innovator groups were assembled, with an average of three participants per group. Throughout the training and pilot phase, participants developed their Gen AI use skills, prompt design, and custom Gen AI model development.

[17] Action research is inherently participatory and iterative, emphasising close collaboration between researchers and practitioners to address practical challenges while generating theoretical and transferable knowledge. Somekh, B. (2005). Action research. McGraw-Hill Education (UK).

[18] A detailed break-down of the survey responses is provided in Appendix 1 of the report.

[19] Rogers, E. M. (2003). Diffusion of Innovations, 5th Edition. Free Press.

Gen AI strategy and model choice

In this project, participants' Gen AI use was undertaken via OpenAI's ChatGPT (including the "custom GPT" functionality). The decision to use ChatGPT followed an internal options appraisal led by the firm's Technology Lead. They drew on knowledge of a range of Gen AI tools to assess suitability for the firm (including functionality, reliability, governance, and data protection considerations) and recommended ChatGPT as the most appropriate platform for the study. Throughout the report, we refer to the technology generically as "Gen AI". The specific tool set up used for each use case (e.g., model selection and settings) is documented within the corresponding case study.

It is important to note that ChatGPT evolved rapidly during the study period in 2025, with frequent updates and major model releases/rollouts (e.g., o3-mini in January 2025; GPT-4.5 in February 2025; GPT-4.1 in April 2025; o3 and o4-mini in April 2025; and the rollout of GPT-5 from August 2025), meaning that capabilities and behaviours can differ across versions.

Following the firm's training, participants were provided access to an enterprise workspace version of ChatGPT. Under OpenAI's enterprise privacy commitments, business data (including prompts and outputs) is not used to train models by default and is subject to enhanced organisational data protections.

Identifying distinctive use cases

The second activity set focused on identifying and investigating the most distinctive and insightful Gen AI use cases among the innovators, which were most likely to provide in-depth insight into the impact of Gen AI on professional judgement. This activity was led by the academic team and involved semi-structured interviews with fifty staff members across various roles and levels of seniority, the review of interview participants' Gen AI-generated outputs and usage logs, and the observation of participants at the B&S Manchester office for up to three full working days each week throughout the project. We used the ICAS Professional Judgement Framework to guide our data collection and initial analysis, noting emerging themes, patterns, and practices related to the impact of Gen AI on professional judgement.

Given the exploratory nature of this research, we were not necessarily interested in the most common Gen AI use cases but in those that would offer genuinely distinctive perspectives on the impact of Gen AI on professional judgement in accounting. To identify a sample, we used the 'knowledge worker' classification developed by Davenport (2005)[21], which differentiates between knowledge work types based on the complexity of interpretation required and the level of interdependence required to achieve the work tasks.[21]

[20] Parker, L. D., & Northcott, D. (2016). Qualitative generalising in accounting research: concepts and strategies. Accounting, Auditing & Accountability Journal, 29(6), 1100-1131. https://doi.org/10.1108/aaaj-04-2015-2026

[21] Davenport, T. (2005). Thinking for a living: How to get better performance and results from knowledge workers. Harvard Business School Press. We also considered the conceptual refinements proposed by Margaryan, A., Milligan, C., & Littlejohn, A. (2011). Validation of Davenport's classification structure of knowledge-intensive processes. Journal of Knowledge Management, 15(4), 568-581. https://doi.org/10.1108/13673271111151965

- **The complexity** of interpretation and judgement in performing a work task determines whether it is foundational, routine, and suitable for junior staff (low complexity), or requires senior professional expertise and judgement (high complexity).

- **The interdependence** dimension captures the extent to which individuals need to rely on other people, processes, or tools in completing the task (low/high interdependence).[22]

The interplay between these two dimensions results in four distinct types of knowledge work:

- **Transaction work.** Routine work with low interdependence and low complexity, typically governed by formal rules and procedures and carried out with limited discretion.

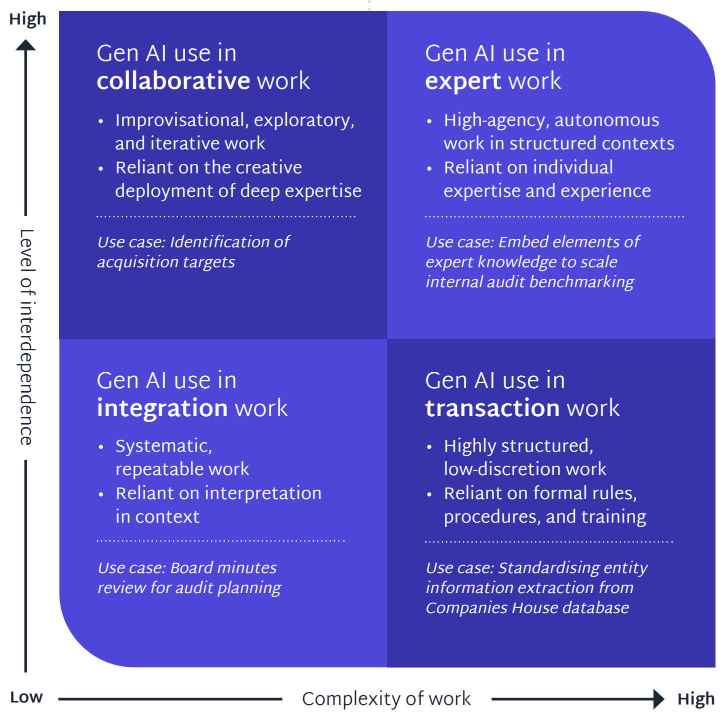
- **Integration work.** Systematic, repeatable work with high interdependence and low complexity, relying on standard processes or methods to coordinate activity across boundaries.

- **Expert work.** Judgement-oriented work with low interdependence and high complexity, relying primarily on individual expertise and experience.

- **Collaboration work.** Learning-oriented work with high interdependence and high complexity, relying on flexible coordination across multiple areas of expertise.

The use cases discussed in this report are drawn from the four resulting quadrants, as shown in Figure 2 below. It is important to note that whilst the four-quadrant model is useful for categorising different tasks and types of use, it is necessarily a simplification: in practice, a single use case may show features of more than one quadrant and may shift category depending on the specific sub-task or the stage of work.

Figure 2: Areas of professional work where Gen AI has been applied



[22] Donoghue, L. P., Harris, J. G., & Weitzman, B. A. (1999). Knowledge management strategies that create value. Outlook, (1), 48–53.

Discussion of the key findings

This report investigated two questions in relation to accounting practices:

1. How does Gen AI impact the professional judgement process of accounting professionals?

2. What are the ethical implications of using Gen AI when developing and making professional judgements in accounting settings?

In response to the first question, the matrix in Figure 2 provides a useful way to interpret the case studies by linking Gen AI use to the level of interpretation needed and the level of interdependence involved in completing the work. The case studies, which are detailed in Appendix 2, suggest that relatively routine and structured tasks (transactional work) may be better suited to standardised uses of Gen AI, such as extraction, summarisation, or the initial organisation of information. By contrast, tasks that require greater contextual interpretation, judgement under uncertainty, or coordination across people and processes (e.g., expert and collaborative work) generally require a clearer outline of Gen AI's role, stronger controls on outputs, and even greater professional oversight of any evaluative conclusions produced by Gen AI. Overall, the case studies show that the usefulness of Gen AI depends on the type of task. A structured assessment of the task's characteristics can help professionals decide when and how to use Gen AI in a way that aligns with the engagement's evidence needs and risk profile.

Across the case studies, we also identify three recurring ways in which professionals remain actively involved and exercise human judgement, when working with Gen AI:

- **Training** involves giving instructions, context, and materials that shape how Gen AI responds; here, we use the term 'training' to mean how users guide the model during use through prompts, constraints, and extra information rather than the machine-learning meaning of adjusting a model's parameters using data.

- **Supervision** requires validating, cross-checking, and critically evaluating Gen AI's outputs.

- **Simulated dialogue** combines training and supervision with some back-and-forth prompting and exploration.

Table 1 illustrates the three modes of human and Gen AI interactions based on the case studies. In practice, we see that the effective use of Gen AI typically involves a combination of all three modes.

Table 1: Examples of the three modes of interaction

	Training	Supervision	Simulated dialogue
Case study 1 **Collaborative work** Daniel's use of Gen AI in the identification of acquisition targets	Daniel set up a custom Gen AI model with defined geography, sector filters, required fields (e.g., ownership and financial indicators), and an Excel-style output format to structure the longlist.	Daniel checked the longlist and flagged gaps or "red flags" by triangulating against trusted sources such as Companies House before using the results.	Daniel used iterative back-and-forth prompting to test alternative sector definitions, refine candidate lists, and reconsider his A/B/C tiering by "bouncing ideas" off the model.
Case study 2 **Expert work** Ahmed's use of Gen AI in internal audit benchmarking	Ahmed configured a custom Gen AI model by providing the internal audit manual, defining the benchmarking areas (e.g., escalation, training, response times), and setting decision rules for "Yes/No/Partial" outputs.	Ahmed spot-checked the table outputs against the underlying policy text, reviewed borderline classifications, and adjusted prompts or criteria when outputs were inconsistent or misleading.	Ahmed rewrote prompts and tested outputs to explore how different instructions affected the benchmarking results.
Case study 3 **Integration work** Henry and Ajay's use of Gen AI in board minutes review for audit planning	Henry and Ajay refined prompts to define what counts as "relevant" (e.g., decisions, actions, risks, attendees, one-off transactions, conflicts) and to produce a standardised table for audit planning.	They spot-checked the Gen AI table back to the underlying minutes and reviewed expected high-risk sections to identify missed or unclear matters and to correct any misclassification.	They used iterative prompting to probe the minutes from different angles (e.g., risks, transactions, conflicts) and to refine the structure and consistency of the extracted issues for the audit team's review.
Case study 4 **Transaction work** Nathan's use of Gen AI to automate pulling out key client information from Companies House	Nathan designed a custom Gen AI model with structured prompts to pull specified Companies House fields (e.g., directors, shareholdings, charges, filings) and present them in a standardised table, while flagging potential risk (e.g., director changes, new borrowing).	Nathan tested the output against the underlying Companies House filings to identify gaps, misidentifications, or hallucinated details.	Nathan asked the model about missing items and inconsistencies and used follow-up prompts to examine specific fields or filings.

In response to the second question, the case studies indicate that ethical considerations arise throughout the professional judgement process, including decisions about whether to use Gen AI for a given task, how confidentiality and data protection are maintained, how outputs are validated and documented, and when additional review is appropriate. On this basis, ethics should be treated as a constant consideration across all stages of professional judgement, rather than addressed only at the point of final decision-making.

All things considered, interacting with Gen AI can introduce efficiencies but also amplify risks and biases.

Efficiencies

- Faster data extraction, summarisation, and cross-referencing of routine information.
- Faster spotting of potential issues when handling large amounts of information and potentially more consistent documentation.
- Automating routine tasks frees up more time for higher-value judgment.

Key risks and biases to manage

- **Automation bias:** accepting coherent Gen AI outputs too readily because they appear authoritative.
- **Confirmation bias:** phrasing prompts in ways that encourage evidence that supports existing expectations.
- **Anchoring bias:** treating Gen AI summaries or risk lists as the baseline, narrowing later scrutiny.
- **Affinity bias:** as with human colleagues or clients, objectivity can be compromised as professionals grow to like or favour the Gen AI they work with.
- **Over-trust through human-like treatment (anthropomorphising):** assuming Gen AI has a real understanding or judgment, which can create misplaced confidence and reduce healthy scepticism.
- **Data privacy and confidentiality:** sharing client-sensitive, personal, or proprietary information in prompts or uploads can create confidentiality breaches and non-compliance risks.
- **Insufficient scrutiny:** all output of a Gen AI process must be checked and verified for accuracy. Spot-checking the output is insufficient for identifying all gaps and errors that might have been made by a Gen AI model.

Spotlight: Gen AI introduces unsolicited output as part of its operations

Across the case studies, Gen AI did not simply condense text; summaries often contained unwanted and unrequested additions. For example, a prompt requesting a neutral overview of internal audit findings might produce statements suggesting that controls were “robust”, “weak in comparison with peers”, or “largely adequate”.

This creates an additional challenge when multiple layers of Gen AI processing are involved. In several cases, judgement was exercised not on the underlying document, but on Gen AI summaries of that document. When professionals reviewed board minutes or internal audit reports, they often reacted to the Gen AI-generated extracts rather than returning to the full source text. While this saved time, it also meant that the scope of judgement was constrained by the Gen AI output. If important nuances and contextual elements were not included in the Gen AI summary, they were unlikely to be considered in the subsequent evaluation.

The risk is not just that Gen AI leaves out important information, but that professionals may unknowingly base their decisions on a filtered version of the evidence, weakening the depth and independence of judgement.

Gen AI is an individualising technology

Professionals and Gen AI build their interaction through repeated dialogue where users refine prompts, correct mistakes, and gradually shape what the system treats as relevant evidence, suitable wording, or plausible reasoning. Over time, these feedback loops can make Gen AI feel like a companion.[23] As we observed in the case studies, practitioners gradually shape how Gen AI searches, summarises, and prioritises information. When Daniel (see Appendix 2) developed “Origination GPT,” or Ahmed rewrote his benchmarking prompts across multiple versions, they were feeding their own judgement and expert knowledge into the system. In this sense, Gen AI begins to internalise aspects of the user’s professional style. Gen AI doesn’t operate as a static tool but as something that is continually evolving and being shaped over time.

This individualised quality has both positive and negative effects on professional judgement. On the positive side, Gen AI personalisation can amplify professional capability: Gen AI can accelerate tasks in ways that feel aligned with how the professional already thinks. The act of training Gen AI can also be reflexive, forcing professionals to articulate what good judgement looks like. Yet the same mechanism risks reproducing and hardening individual bias. If Gen AI gradually adapts to a user’s prior assumptions, preferred risk cues, or usual ways of interpreting things, Gen AI can lead to a confirmation bias, subtly confirming what professionals already think instead of challenging it. For example, if a user repeatedly asks Gen AI to look for governance risks based on provided board minutes, the model may keep highlighting similar issues each time and overlook other plausible risk factors.

Finally, AI’s individualised behaviour makes visibility and oversight more difficult. If each professional is effectively working with a slightly different Gen AI model, then keeping judgement consistent across teams becomes difficult. This highlights a further organisational risk: Gen AI personalisation can drift into informal or unreported use, especially when staff worry about peer judgement, uncertainty policies, or the stigma of relying on Gen AI. Hidden, individualised Gen AI adoption makes it difficult for firms to spot automation bias or learn collectively from how Gen AI is being used.

In our case studies, we did not observe Gen AI being used in a group or team setting (for example, jointly prompting or collectively reviewing outputs in real time). However, Gen AI output from one task can easily become the input for another workflow that also uses Gen AI. For example, internal auditors may use Gen AI to draft or summarise internal audit findings, which are then relied on by external auditors as part of assessing the internal audit function and deciding on planned audit procedures. If external auditors then use Gen AI again to summarise, interpret, or integrate that internal audit work into their own planning records, this creates a layered chain of Gen AI interactions. We refer to these as feedback loops, in effect, “Gen AI talking to Gen AI”. This presents an additional risk to manage and safeguard against, because successive layers of Gen AI processing can amplify bias and errors.

Gen AI puts professional ethics in the spotlight

The ICAS (2016) Professional Judgement Framework positions ethics as central to sound decision making. It emphasises the importance of tone at the top and the need for clear escalation and approval procedures so that key judgements are made at the right level. ICAS also advises practitioners to draw on the ICAS Ethical Decision-Making Framework, the ICAS Code of ethics derived from the IESBA (International Ethics Standards Board for Accountants) Code[24], and consultation with colleagues when dilemmas arise.

Amid rapid technological and digital change, both IESBA and subsequently ICAS have revised their respective Codes of Ethics regarding the role of various technologies. For instance, these revisions include the need to identify the risks that new technologies create for accountants and auditors, and to use professional judgement to decide what actions, if any, are required to guard against bias or becoming linked to misleading information. Gen AI raises a range of ethical concerns that have sparked significant debate in both academic and practitioner circles. For example, Hagendroff (2024) identified 18 categories of ethical topics that relate to Gen AI, consisting of fairness-bias, safety, harmful content-toxicity, hallucinations, privacy, interaction risks, security-robustness, education-learning, alignment, cybercrime, governance-regulation, labour displacement-economic impact, transparency-explainability, evaluation-auditing, sustainability, art-creativity, copyright-authorship, and writing.[25]

[23] Research in human–computer interaction shows that people often respond to interactive technologies as social partners, especially when those technologies communicate using language and other human-like cues. Recent work argues that large language model chatbots can mimic human conversation convincingly, which increases the likelihood that users frame the interaction in interpersonal terms. See, e.g., Peter, S., Riemer, K., & West, J. D. (2025). The benefits and dangers of anthropomorphic conversational agents. Proceedings of the National Academy of Sciences, 122(22), e2415898122.

[24] ICAS. Code of Ethics. 2025, p.17.

[25] Hagendorff, Thilo. ‘Mapping the Ethics of Generative AI: A Comprehensive Scoping Review’. Minds and Machines 34, no. 4 (2024): 39. https://doi.org/10.1007/s11023-024-09694-w

From a user perspective, the following extracts from the fundamental ethics principles, contained in the IESBA Code and ICAS Code of Ethics, are particularly relevant to Gen AI use:

- **Objectivity:** Exercising professional and business judgement free from undue influence or reliance on technology or other factors.
- **Professional competence and due care:** Maintaining professional knowledge and skill as well as acting diligently and in accordance with applicable technical and professional standards.
- **Confidentiality:** Respecting the confidentiality of information acquired through professional and business relationships.

In this context, our findings show that ethics should be built into each step of professional judgement In Gen AI-supported workflows, this has two immediate implications.

Firstly, professionals may need the moral courage to challenge or refuse the use of Gen AI when it would undermine the fundamental principles, for example, when the outputs cannot be properly checked, when confidentiality could be breached, or when there is pressure to accept an easy answer. ICAS describes moral courage as a core quality that helps professionals follow the fundamental ethical principles of professional conduct, particularly when doing the right thing could come at a personal cost.

Secondly, the ethical use of Gen AI depends on a solid understanding of its strengths and limitations, so that risks such as bias, misleading results, or misuse of data can be identified and addressed through appropriate safeguards and controls. Across both points, we stress that the professional is still accountable for the reasoning, decisions, and outcomes. Gen AI can help with different stages of professional judgement, but it does not take on professional duties or responsibility for outcomes, and it cannot replace the practitioner's duty to use professional judgement and scepticism.

Overall, these findings point to several ethical areas that need attention:

- Safeguards against automation bias and uncritical acceptance of Gen AI outputs
- Expectations for documenting how Gen AI was used in gathering or analysing evidence.
- Having solid escalation processes for when Gen AI produces results that look incomplete and inconsistent.
- Protecting practical learning so that junior staff keep developing the interpretive skills needed for ethical judgements.

Future areas of focus for the profession

Drawing on our research, we highlight the key areas the accountancy profession should explore and act on in the future.

Embedding chartered accountancy within Gen AI systems

Through careful prompts and repeated correction, participants tried to teach Gen AI the norms of professional conduct, scepticism, and analytical reasoning. This raises key questions about whether professional bodies such as ICAS will need frameworks for training GPTs, and if training Gen AI systems will become as important as human education to maintain high levels of trust in the profession.[26]

Preserving time for reflection in professional judgement

Gen AI produces instant answers, cutting out the time usually given for reflective and deliberate thought. While this speed can improve efficiency, it also risks undermining one of the cornerstones of professional judgement: the capacity to pause, reflect, and engage in reasoned scepticism. These are professional qualities that immediate AI responses cannot replicate. Auditing standards already recognise the importance of such reflection through the "stand-back" requirements in the International Standards on Auditing (ISAs), which require auditors to step back and check whether the evidence gathered is still sufficient and appropriate, including looking for anything that contradicts it. Future research should examine how organisations can design workflows that build in time to weigh evidence even when using AI-driven processes.

Preparing for the future of accountability, transparency, and regulation

As Gen AI begins interacting with its own outputs and those of others, it will become harder to know what decisions are human-led and what comes from AI. While AI use in accounting is still emerging, professional judgement requires the human professional to always remain responsible and accountable for the decision-making in AI-mediated processes.[27] This poses a systemic risk to the integrity of professional work, which should be considered in future regulation and professional guidance to ensure that human responsibility remains traceable and auditable.

[26] ICAS. (2025). Society First: Shaping the profession. https://icas-com.uksouth01.umbraco.io/media/j3zl4laz/ic0008_stp_report_longform_v3.pdf
[27] Accounting Web (2025) article on Agentic AI. https://www.accountingweb.co.uk/tech/tech-pulse/agentic-ai-edges-closer-to-accounting-adoption

Maintaining broad access to the profession

An additional consideration for the profession is the potential for Gen AI to influence or exacerbate inequality and access within the profession. Although our concept-driven selection of Gen AI use cases among innovator group participants mainly highlighted cases developed by male staff, which may be coincidental, recent research suggests that Gen AI tools may reinforce gender disparities and widen existing skills gaps.[28] Organisations should make sure that all staff members, regardless of gender or other characteristics, receive training and support.

Turning Gen AI into the engine for professional growth, training, and education

Finally, the case studies in this report show that the fast growing use of Gen AI in accounting will greatly increase the need for skilled and experienced professionals who can confidently challenge its outputs.

Yet, a common concern in the accounting profession is that Gen AI might reduce junior staff's practical learning opportunities by taking over their tasks or narrowing the experience gap between junior and senior staff, affecting career progression and the growth of sound professional judgement.[29]

Nevertheless, there are also opportunities for Gen AI to challenge these concerns. While Gen AI may disrupt traditional experiential learning, it does not necessarily lead to a loss of expertise or professional standards. Instead, the profession can play an active role in shaping how Gen AI is used as a learning environment where expertise continues to develop alongside the technology. For example, there might be useful insights from educators who have successfully integrated Gen AI into teaching to support personal growth, knowledge, and skill development.[30]

ICAS is well-positioned to lead the profession in addressing these challenges and ensuring that technology enhances, rather than diminishes, the quality and value of professional judgement.

[28] Otis, N. G., Delecourt, S., Cranney, K., & Koning, R. (2025). Global Evidence on Gender Gaps and Generative AI (25-023). Harvard Business School. https://www.hbs.edu/ris/Publication%20Files/25023_52957d6c-0378-4796-99fa-aab684b3b2f8.pdf ; Roldan-Mones, A. (2025). When Gen AI increases inequality: Evidence from a university debating competition. London School of Economics and Political Science. https://poid.lse.ac.uk/textonly/publications/downloads/poidwp096rev.pdf .
[29] Krakowski, S. (2025). "Human-Ai Agency in the Age of Generative Ai." Information and Organization 35(1): 100560.
[30] See, e.g., https://www.timeshighereducation.com/campus/chatgpt-teaching-tool-not-cheating-tool.

Appendix 1. A selection of survey responses

Table A1: Responses per department

Department	Responded	Total	Response %
Administration	8	16	50%
Compliance	2	2	100%
Corporate Finance	3	4	75%
Corporate Recovery	0	2	0%
Digital Transformation	7	9	78%
External Audit	101	160	63%
Freelancers	5	5	100%
HR	4	4	100%
Marketing	4	5	80%
Payroll	5	8	63%
Internal Audit	34	42	81%
SME	18	19	95%
Tax	10	16	63%
Total	201	292	69%

Table A2: Responses per role

Role	Responded	Total	Response %
Partners	12	16	75%
Directors	10	13	77%
Managers	46	62	74%
Trainees	133	201	66%
Total	201	292	69%

Table A3: Which tasks have you used Gen AI for?

Categories provided in survey	Ranking	Quantity	Percentage
Writing emails and editing text	1	63	53%
Generating summaries (meetings, documents)	2	55	46%
Studying for the exams	3	52	44%
Working with Excel (or other software)	4	39	33%
Solving quantitative problems	5	25	21%

Table A4: What concerns, if any, do you have about AI in your work?

Categories provided in survey	Ranking	Quantity	Percentage
ChatGPT making errors or incorrect decisions	1	140	72%
Lack of training or understanding	2	105	54%
Client data privacy	3	101	52%
Job security	4	41	21%

Table A5: How did Gen AI change the pace at which you perform your tasks?

	Quantity	Percentage
Increased a lot	27	22%
Increased a little	63	52%
No effect	27	22%
Decreased a little	1	1%
Don't know	3	2%

Appendix 2. Case studies of human and Gen AI interactions in the context of professional judgement

Case study 1 | Collaborative work

Daniel's use of Gen AI in the identification of acquisition targets

Role and context

Daniel is a Corporate Finance Executive with four years of industry experience. His work involves mapping potential acquirers and sellers of businesses within a sector, classifying them into priority tiers (e.g., A, B, C), and advising clients on acquisition opportunities. He began experimenting with Gen AI during the B&S pilot trial. Building on these early successes, Daniel participated in the B&S innovators' workshop to design a custom Gen AI model that he named "Origination GPT."

Type of work

This task aligns with collaborative work because it is high complexity and high interdependence: the classifications depend on combining financial indicators with wider deal knowledge, such as ownership structures, strategy, governance, management capabilities, and potential strategic or cultural fit.

Gen AI setup

Tool	ChatGPT
Model	Version 4o
Web search	Enabled

Before Gen AI

Prior to the Gen AI trial, Daniel used a combination of Google searches, LinkedIn checks on personnel involved, and specialist databases such as Mark to Market and Market IQ for knowledge gathering. This was a labour-intensive process which could take days or weeks, depending on the client requirements.

With Gen AI

With the introduction of Gen AI, Daniel found he could widen the scope of his initial searches and then use Gen AI to "narrow down lists" and "fill in gaps." The lists generated were not definitive, but they enabled him to double-check classifications, confirm prior judgements, and consider alternatives he might otherwise have overlooked.

The custom Gen AI model was programmed to generate business lists filtered by geography, sector, and key financial or governance indicators. More specifically, "Origination GPT" generated an Excel-style longlist of candidate firms and pre-populated key fields (e.g., location, sector, short description, and selected financial indicators), while also flagging gaps and potential "red flags" to prompt Daniel's verification against sources such as Mark to Market, Market IQ, and Companies House. Such longlists were intended to focus primarily on UK-registered companies, with the practical emphasis on the North West of England (with possible extension to nearby regions) rather than a UK-wide scan.

In this context, the ability to create a custom Gen AI model allowed Daniel to define the search scope, the categories to be captured, and the format of outputs across both accounting and non-accounting information, speeding up early-stage exploration while retaining the need for verification and professional judgement.

<table>
<tr><th colspan="2">Impact on the professional judgement process</th></tr>
<tr><td>Principle 1
Knowledge gathering and analysis</td><td>Daniel used "Origination GPT" mainly to speed up desk research for a chosen geography. The tool was designed to follow a partly structured workflow that required explicit human input at predefined stages:
(1) user input on which area to focus on,
(2) user input on which sectors to include,
(3) Gen AI producing a draft table of potential businesses with the requested fields (e.g., business name, location, and brief business details, as well as financial and ownership fields where available), and
(4) Gen AI prompted Daniel to check the results and then compare and complete the information using other databases.
This design illustrates that even "exploratory" Gen AI use involves process design by the accounting professional, including structuring the sequence of steps and embedding requirements for human judgement and verification before outputs are used in practice. Daniel referred to the process metaphorically as "bouncing ideas off Gen AI" and "using [Gen AI] as a sounding board".
For Daniel, the expected benefits of the custom Gen AI model materialised through this combination of a partly structured workflow and rapid text generation. Daniel perceived his interactions with Gen AI as an iterative dialogue that helped him test sector boundaries and revisit prior assumptions. In this sense, he experienced the custom Gen AI model as an enhancement of his professional judgement process at the intersection of multiple types of knowledge, including sector understanding, ownership and governance considerations, and financial information.</td></tr>
</table>

<table>
<tr><th colspan="2">Impact on the professional judgement process</th></tr>
<tr><td>Principle 2
Assessment of accounting guidance</td><td>The professional judgement challenges did not centre on applying accounting standards or other authoritative accounting literature; instead, they arose in a domain where accountants must work beyond standard accounting guidance and integrate multiple forms of knowledge. The task required understanding legal requirements and relevant financial regulation, alongside sector and deal context, and translating these inputs into advice that depends on professional interpretation by the accountant in areas that are often relatively unstructured and less codified. In this setting, Gen AI was helpful primarily as a support for bringing together and organising interdisciplinary information to inform the judgement process.</td></tr>
<tr><td>Principle 3
Process for making a judgement</td><td>Daniel's judgement process defined the evaluative criteria that would inform judgements of acquisition opportunities for his clients. Even when Gen AI generated knowledge that Daniel might not have otherwise considered, he maintained control over the evaluation and verification of these outputs.</td></tr>
<tr><td>Principle 4
Documentation of judgement</td><td>In this case study, we observed little change in the documentation of judgement itself: Daniel still needed to record his rationale, key assumptions, and the basis for any final recommendations to the client.</td></tr>
<tr><td>Reflections on the role of the accounting professional</td><td>Daniel's workflow combined two elements: some outputs were shaped in advance through the custom Gen AI model setup choices, which placed him in a role of supervising whether the tool was functioning as intended. At the same time, once the tool produced an initial longlist by grouping semantically related information, Daniel engaged in simulated dialogue through iterative prompting to narrow options, test his A/B/C tiers, and identify additional targets. The case study points to two good practices for collaborative tasks that rely on broader business knowledge: customise the Gen AI model with clear context and constraints at the outset and then supervise its outputs by checking them against trusted sources and further refining them.</td></tr>
</table>

Case study 2 | Expert work

Ahmed's use of Gen AI in internal audit benchmarking

Role and context

Ahmed is a qualified Chartered Accountant working in the Internal Audit function at B&S. He has three years' experience at the firm, following prior experience at another accounting practice in a similar role. Ahmed developed a custom Gen AI model to streamline the internal audit benchmarking process, and, according to him, to further standardise it. The ambition of technology is to "embed" elements of expert knowledge into a repeatable assessment process.

Type of work

This case reflects Gen AI use in an expert-work context, where the ambition of technology is to "embed" elements of expert knowledge into a repeatable assessment process. Internal audit benchmarking requires auditors to interpret structured compliance frameworks and apply them in nuanced organisational contexts. The task scores relatively low on the interdependence dimension because it can usually be completed by an individual auditor using documented policy requirements and established control expectations, without extensive coordination across colleagues or teams. Yet it is also complex because each organisation's policy articulates these requirements differently, demanding professional interpretation and judgement. That interpretation depends on the auditor's personal professional knowledge and experience, including their understanding of what constitutes sufficient evidence and adequate articulation in light of the relevant control expectations and the specific organisational context.

Gen AI setup

Tool	ChatGPT (Enterprise)
Model	Version 4o
Web search	Enabled

Before Gen AI

Historically, benchmarking involved manually comparing client policies (e.g., fire safety, gas inspections) against peers, either by reusing an old data table or building a new one. According to Ahmed, both methods were labour-intensive and inconsistent because they relied heavily on the auditor's personal expertise to interpret how policy wording mapped onto the benchmarking criteria, which made results sensitive to individual judgement.

This is characteristic of expert work, where key interpretive knowledge is located "in the brain" of experienced professionals, and the practical challenge is to embed that expertise into a more repeatable workflow without losing necessary professional judgement.

With Gen AI

Ahmed developed a custom Gen AI model to streamline the internal audit benchmarking process, and, according to him, to further standardise it. Using Ahmed's custom Gen AI model, auditors could upload a client policy, and Gen AI would test it against pre-defined compliance areas, such as escalation procedures, training protocols, and target response times, before returning a table indicating "Yes," "No," or "Partial" coverage, complete with extracts and references.

As Ahmed explained, the custom Gen AI "tests the benchmarking areas with the policy ... and it basically does the test." According to Ahmed, this output could then be integrated into audit working papers or a cumulative benchmarking table across clients. He also experimented with what he described as a "perfect policy," a dynamically updated model that "integrates best practice from all policies reviewed".

The utility of Gen AI in ensuring a higher level of standardisation remained unclear. This is because instead of reliably repeating the same output from the same input, Gen AI produces new, probabilistic responses each time and occasionally introduces details, interpretations, or omissions that were not present in the original material. Such variability makes standardisation difficult to guarantee.

Impact on the professional judgement process	
Principle 1 Knowledge gathering and analysis	In this first stage, Gen AI reduced the time needed for benchmarking. Ahmed said that before using the custom Gen AI model, benchmarking one policy could take three to four hours because auditors had to read the full document, pull out the relevant requirements, and manually build the comparison table. With the custom Gen AI model, Ahmed could source peer internal audit policy documents from the internet in minutes, upload a policy, and have it summarised and converted to a benchmarking table in half an hour or less. This also changed Ahmed's workflow and areas of attention. Instead of spending most of his time reading source documents and extracting points himself, he increasingly started with Gen AI summaries and tables. He focused on checking accuracy, reviewing any "Yes/No/Partial" decisions, and following up on unclear or borderline areas in the original policy text. This illustrates Gen AI's potential in expert work: it can shift effort from completing the full task manually to supervising defined outputs and specific stages of the tasks, allowing the expert to focus on the parts of the workflow that most require professional judgement.

Impact on the professional judgement process	
Principle 2 Assessment of accounting guidance	In the second stage, the assessment of accounting or compliance guidance, Gen AI subtly influenced how the criteria were applied. First, Ahmed provided Gen AI with a firm's internal audit manual to perform the task: "I have provided [custom Gen AI model] with our audit manual. It will ask you three questions: 1) Which policy do you want to benchmark? What type of organisation is the client? Approximate size of the organisation? So, let us say "Water safety, housing". It will search the internet and find three comparable water safety policies [...]. This is not the first version of the tool. This is the fourth version where I had to rewrite the prompt from scratch [...]. It will now create a benchmarking table out of these policies. " Gen AI was tasked with filtering and organising evidence, drawing Ahmed's attention to potential weaknesses or missing elements. By filtering through large, semi-structured documents and subsequently turning them into formatted markdown tables aligned with benchmarking areas, Gen AI guided Ahmed's focus and reasoning process, influencing what he perceived as key strengths and risks.
Principle 3 Process for making a judgement	In the third stage, Ahmed's reasoning was shaped by the Gen AI output, which could introduce potential for bias. Instead of beginning with the raw policy and forming his own view about what was relevant or how well the client's policy aligned with expectations or accepted practice, he evaluated the classifications, summaries, and extracts that Gen AI had already highlighted. His task became checking whether the Gen AI output fairly presented the underlying policy, in addition to checking whether the suggested coverage was reliable, and whether anything in the source material contradicted the proposed assessment. Gen AI's organisation of the material influenced which areas appeared most significant and which comparisons seemed appropriate, meaning that Ahmed's judgement was more focused on confirming or challenging the structure that Gen AI had produced.
Principle 4 Documentation of judgement	The Gen AI model transformed the documentation of judgement. Previously, auditors manually built benchmarking tables, recording each judgement as they went. This act of manually writing up the judgements formed part of the reasoning process. With the introduction of the custom Gen AI model, documentation became instantaneous. Gen AI provided output in a markdown-formatted table containing compliance ratings, textual extracts, and references to policy sections.
Reflections on the role of the accounting professional	Ahmed's experience illustrates a paradox. The low interdependence of audit benchmarking made it promising for Gen AI integration, but the act of codification exposed where expertise still mattered. Ahmed had to translate internalised professional expectations into explicit benchmarking criteria, define how the tool should interpret common policy language, and set decision rules for what should count as "Yes," "No," or "Partial." He then supervised outputs by checking cited extracts against the source policy, reviewing borderline classifications, and revising the prompts and criteria when results were inconsistent or misleading.

Case study 3 | Integration work

Henry and Ajay's use of Gen AI in board minutes review for audit planning

Role and context

Henry and Ajay are junior auditors in training to be Chartered Accountants. Henry developed a custom Gen AI model to streamline the review of board minutes in the context of audit planning. Ajay collaborated with Henry to refine the prompts, emphasising that board minutes reviews should follow a standardised format.

Type of work

Henry and Ajay's work is located in the integration quadrant: it is relatively low in complexity because it relies on established audit concepts and a standard approach to identifying potential risks, but it is relatively high in interdependence because outputs must align with the firm's audit methodology and feed into team planning, review and documentation, requiring coordination across the audit team and contextual awareness. The underlying objective of the task is to support the identification of risks of material misstatement by scanning minutes for relevant topics and signals that the audit team should consider.

Gen AI setup

Tool	ChatGPT (Enterprise)
Model	Version 4o
Web search	Disabled

Before Gen AI

Board minutes reviews are a standard element of audit planning and fieldwork. The purpose of this task is to identify matters recorded in client board meetings that may have an impact on the audit, such as acquisitions, litigation, or changes in senior management. Board minutes can also shape certain areas of the audit if information comes to light that could lead to further or new testing. Because board minutes are produced primarily for internal governance purposes, they typically contain large volumes of information that are non-essential and discretionary to the external audit. Consequently, reviewing them can involve hundreds of pages of documentation, with only a handful of items proving material for audit evidence or risk assessment. In practice, this task is often assigned to junior auditors, serving as a labour-intensive screening activity for developing professional judgement and contextual awareness.

With Gen AI

A custom Gen AI model was designed to scan uploaded board minutes, extract relevant information into a structured table, and highlight potential audit-relevant issues. Consistent with the opportunities for technology in integration work, the custom model was intended to recognise recurring language patterns in minutes that may indicate potential risks and then present these in a consistent output for the auditor to assess. In other words, Gen AI was used to improve coverage and consistency in identifying potential risks, while leaving the accounting professional to judge whether the issues identified constitute a risk of material misstatement and how they should be addressed in the audit plan.

To guide what was treated as "relevant", Henry provided prompts such as: What decisions were made in the last board meeting?; Summarise the actions assigned in the latest minutes.; Are there any risks mentioned in this board document?; What was the date of the board meeting, and who was present at the meeting?; Are there any large one-off transactions (provisions, fees, expenses, penalties) included in the board minutes? And are there any conflicts of interest declared in the meeting?

In this case, auditors checked for missed matters by spot-checking: they traced a sample of items in the Gen AI table back to the underlying minutes and reviewed sections they expected to contain key audit issues, following up on anything unclear in the source document. Spot-checking is insufficient for identifying gaps and errors made by Gen AI, and at present, all Gen AI output should be checked against source material.

Impact on the professional judgement process	
Principle 1 Knowledge gathering and analysis	In the knowledge-gathering stage, Gen AI altered the nature of the professional workflow. Traditionally, auditors read board minutes, identifying indicators of risk or audit relevance through immersion in the narrative context. The custom Gen AI model replaced this with rapid extraction and classification, producing structured tables of potential issues within minutes. What once took several hours of human scanning could now be completed in less than half an hour. The act of reading was substituted by a review of summaries, shifting the auditor's engagement from exploration to verification of Gen AI output.
Principle 2 Assessment of accounting guidance	In this case study, we documented the limited impact of Gen AI on the assessment of accounting guidance within the professional judgement process, largely because the custom Gen AI model was used primarily for the early-stage workflow. The custom Gen AI model supported professional judgement by highlighting parts of board minutes that could indicate audit risks. While the prompts asked Gen AI to flag categories that are commonly relevant to audit risk (for example, large one-off transactions such as provisions, fees, expenses, or penalties), the subsequent steps, namely linking any flagged items to the relevant accounting requirements and judging whether any potential impact could be material, remained with the auditors.

Impact on the professional judgement process	
Principle 3 Process for making a judgement	In the third stage, the process of making judgement, Henry's reasoning was applied to the output from Gen AI, and this created a risk of overreliance because the Gen AI model carried out much of the initial selection and organisation of the information before Henry reviewed it. Instead of identifying relevant issues through his own reading of the board minutes, he started with the custom Gen AI model output of items and summaries, which directed his attention toward what the model had highlighted as potentially significant. Professional judgement was still required, but it was applied mainly to verifying and refining the Gen AI structuring of the material rather than constructing that structure himself.
Principle 4 Documentation of judgement	The documentation stage was likewise impacted. The custom Gen AI model automated much of this work as the output produced structured tables complete with references and draft commentary. As a result, part of the reasoning traditionally developed through engagement with the minutes writing was displaced.
Reflections on the role of the accounting professional	The streamlining of board minutes reviews illustrates how Gen AI can automate the areas of workflow that have traditionally served as apprenticeship spaces for learning professional judgement. Auditors still worked through the material, but instead of preparing the initial summaries themselves, they reviewed and assessed the output produced by the Gen AI model.

Case study 4 | Transaction work

Nathan's use of Gen AI to automate client information extraction from Companies House

Role and context

Nathan is an Audit Manager at B&S with four years of experience at the firm. He designed a custom Gen AI model to automate the extraction of standardised company information from the Companies House database.

Type of work

This workflow aligns closely with Davenport's matrix as a low interdependence and low complexity task: the workflow draws on a specified external database (Companies House), follows predefined retrieval steps, and targets standardised fields. Contingencies are largely predictable (e.g., missing filings, inconsistent naming conventions), making the task amenable to automation through structured prompts and validation checks.

Gen AI setup

Tool	ChatGPT (Enterprise)
Model	Version 5
Web search	Enabled

Before Gen AI

As part of audit planning under ISA 315, auditors are required to obtain an understanding of the entity and its environment. Part of Nathan's work involves gathering information from structured, public sources such as Companies House, identifying directors, shareholders, mortgages, debentures, and recent filings. While the task is procedurally standardised, it still requires discernment to interpret changes in governance or financing that may affect audit risk.

With Gen AI

The custom Gen AI model was prompted to retrieve a company's public filings, extract key details, and summarise them into a structured table. The prompt also instructed the model to highlight relevant observations, such as director changes or new borrowing arrangements that might signal potential risk factors.

At first, the Gen AI output appeared efficient, returning structured summaries within seconds, neatly formatted and ready for inclusion in audit files. However, Nathan realised the results were unreliable and inconsistent as some of the key information was occasionally omitted, directors were misidentified, and/or showed signs of hallucinations with plausible but false details. When Nathan queried the omissions, the Gen AI model sometimes filled the gaps with invented answers.

Impact on the professional judgement process	
Principle 1 Knowledge gathering and analysis	In ISA 315, understanding the client's governance and control environment is not a descriptive task but an evaluative one. The auditor must determine whether factual changes, such as new directorships, complex ownership structures, or debt arrangements, indicate heightened risk. Gen AI accelerated the data collection process. What once took an auditor thirty to forty minutes of manual review per company, reading filings, noting changes, and cross-referencing, could now be achieved in under five minutes. Yet this increased speed transformed the nature of the task as it also decreased the accuracy of the results. Rather than constructing understanding through exploration, Nathan encountered knowledge that was organised and pre-summarised by Gen AI.
Principle 2 Assessment of accounting guidance	In assessing the requirements of ISA 315, Nathan focused on whether the information generated by Gen AI genuinely supported the auditor's obligation to obtain an understanding of the entity and its environment, including governance structures, ownership, financing arrangements, and indicators of potential risk.
Principle 3 Process for making a judgement	Exercising professional judgement required Nathan to move beyond the authoritative appearance of the Gen AI outputs and actively test their reliability against the underlying public filings. His process involved searching for omissions and mismatches using these as signals that further inspection was needed.
Principle 4 Documentation of judgement	The influence of Gen AI was most visible in the documentation stage. Gen AI-generated summaries were formatted in markdown tables with headings and descriptions, ready for direct inclusion in audit files.
Reflections on the role of the accounting professional	Gen AI's role in this case study is to automate the sourcing and extraction of client background information from Companies House, converting public filings into usable summaries and tables for audit work. Nathan's custom Gen AI model also shows broader potential: future custom Gen AI models could connect directly via Companies House APIs[31], further automating these standardised searches. This could reshape the informational infrastructure and interface through which professionals engage with regulatory data, which is necessary for their everyday judgement.

[31] Application Programming Interface

ICAS - The Professional Body of CAS
icas_accounting
@ICASaccounting
icas_accounting