ÁREA TEMÁTICA: Casos de Ensino

CATEGORIA: ENANGRAD PLENO

WHEN ASSURANCE FAILS: TANIA RISHNIW AND THE TARGETED COMPLIANCE FRAMEWORK

Resumo

Este caso de ensino situa-se na governança de inteligência artificial no setor público. Em setembro de 2025, Tania Rishniw, Secretária-Adjunta do Departamento de Emprego e Relações de Trabalho da Austrália, descobre que a Deloitte usou ferramentas de IA generativa não divulgadas para produzir parte de um relatório de garantia de A$440 mil sobre um sistema automatizado de penalidades assistenciais, entregue com vinte citações alucinadas, incluindo uma citação judicial fabricada que sustentava a principal conclusão jurídica. O objetivo do caso é levar o leitor a decidir, sob prazo de quatro dias, a postura institucional da protagonista: quanto exigir de volta, se escalar pelo Código de Conduta de Fornecedores, se publicar o relatório corrigido e se exigir a correção da carta da firma. Metodologicamente, é um caso decisório construído apenas com fontes de registro público — registros parlamentares, documentos de acesso à informação, decisão da Justiça Federal e comunicados —, sem entrevistas, indicado para cursos de pós-graduação e educação executiva. A análise apresenta cinco opções não excludentes e seus custos, distinguindo falhas de processo de falhas de substância e usos de IA divulgados de não divulgados. Conclui-se que, em trabalho profissional regulado, o problema de governança não é o erro do modelo, mas a relação de divulgação, consentimento e garantia de qualidade em torno dele, e que frameworks de ética em IA só governam quando operacionalizados em procedimentos, treinamento e incentivos.

Palavras-chave: governança de inteligência artificial; compras públicas; ética em serviços profissionais; responsabilização algorítmica; caso de ensino.

Abstract

This teaching case addresses artificial intelligence governance in the public sector. In September 2025, Tania Rishniw, Deputy Secretary of Australia's Department of Employment and Workplace Relations, learns that Deloitte used undisclosed generative AI tools to produce part of an A$440,000 assurance report on an automated welfare-penalty system, delivered with twenty hallucinated citations, including a fabricated judicial quotation anchoring its central legal finding. The case asks the reader to decide, within four days, the protagonist's institutional posture: how much money to recover, whether to escalate under the Commonwealth Supplier Code of Conduct, whether to publish the corrected report, and whether to demand correction of the firm's letter. Methodologically, it is a decision-forcing case built entirely from public-record sources — parliamentary Hansards, Freedom of Information releases, a Federal Court judgment, and attributable press releases — with no interviews conducted, suited to graduate and executive courses. The analysis presents five non-exclusive options and their costs, distinguishing process failures from substantive ones and disclosed from undisclosed AI use. It concludes that, in regulated professional work, the governance problem is not the model's error but the disclosure, consent, and quality-assurance relationship around it, and that AI-ethics frameworks govern only

when operationalised into the procedures, training, and incentives that bind engagement teams in real time.

Keywords: artificial intelligence governance; public-sector procurement; professional services ethics; algorithmic accountability; teaching case.

The Case

Case author's note. This case was prepared from materials in the public record: parliamentary Hansards, documents released under Australia's Freedom of Information Act 1982, consent orders of the Federal Court of Australia, Deloitte's Targeted Compliance Framework Final Report, and attributable press releases. No interviews were conducted. Redactions made under section 22 of the FOI Act are respected. The case is intended for classroom discussion and is not a comprehensive account of the events it depicts.

1. The 2 September letter

On the evening of Friday, 5 September 2025, Tania Rishniw closed the door of her office on the fourth floor of 50 Marcus Clarke Street, Canberra, and reread a three-page letter that had reached her that afternoon through the Department of Finance. The letterhead read Deloitte Touche Tohmatsu. The letter, dated 2 September 2025, was addressed to Anush Avakian, First Assistant Secretary, Commercial Division, Department of Finance, not Tania's own department. It had been forwarded to her, three days late, by a Finance officer who noticed that the version of events it described did not match the version her department, the Department of Employment and Workplace Relations (DEWR), was working with.

That morning she had walked the Lake Burley Griffin parkland to clear her head before the week's second meeting with Deloitte's engagement team. Two weeks earlier, a journalist from the Australian Financial Review had asked Tania whether she stood by the citations in the Targeted Compliance Framework Independent Review Final Report, a report her department had paid Deloitte A$440,000 to produce. She had asked Deloitte for an urgent answer. The reply described "incorrect citations because of a transcription issue." She had asked again. The answer had not changed.

Now, in writing, Deloitte was telling another department something different. The letter said Deloitte had used generative artificial intelligence in two stages of the engagement. The first, analysis of the IT code underpinning the Targeted Compliance Framework, DEWR had approved. The second, using generative AI to summarise a Federal Court judgment and to format citations, DEWR had not approved and had not been told about. The letter also argued that the citation problems did not amount to a Significant Event under the head agreement governing the engagement, and asked Finance to treat the matter accordingly.

Tania set the letter down and lined up four decisions. By Tuesday morning she would need to write her own letter to Deloitte, settling each of them: how much of the A$440,000 to demand back; whether to escalate under the Commonwealth Supplier Code of Conduct; whether to publish Deloitte's promised corrected report; and whether to demand that Deloitte correct the letter on her desk. Senate Estimates would convene in five weeks. Whatever she did, or did not do, this weekend would by then be visible (Exhibit A sets out the timeline of events).

2. The mandate Tania had inherited

Tania Rishniw was Deputy Secretary of the Employment and Workforce Group at DEWR, the department responsible for Australia's welfare-conditionality regime, which paid roughly A$1.385 billion in 2024-25 to job-services providers. She had inherited a policy headache that had become a public disaster.

The TCF was the IT system through which the government decided whether unemployed welfare recipients had met their mutual obligations: attending appointments, applying for jobs, attending training. A recorded breach could suspend a payment; repeated suspensions could cancel it. By 2025 the system was generating more than 280,000 suspensions per quarter, touching roughly 280,000 of the 800,000 recipients with mutual-obligation requirements at any time.

The TCF carried a difficult precedent: Robodebt, the 2015 automated debt-recovery scheme that unlawfully averaged income data to demand repayments. In 2019 the Federal Court found, by consent, that one such demand was irrational for want of any probative material to support it. A 2023 Royal Commission found the scheme unlawful and found that senior officials had failed in their duties; the Commonwealth eventually settled a class action for A$1.8 billion. The TCF operated on a different legal footing, yet it remained automated and applied at scale, and the memory of Robodebt was present in every discussion of it. In November 2024 Tania commissioned the TCF's first independent assurance review since 2018, through the Management Advisory Services Panel administered by the Department of Finance. Deloitte's proposal won; the Order for Services issued on 23 December 2024. Her remit was the integrity of the system she ran; the identity of the winning supplier was a market outcome, and its commercial fortunes were not her concern.

3. Deloitte in Canberra in 2025

Deloitte was the largest of the Big Four firms by 2025 revenue. On 30 September, four days after delivering DEWR a corrected TCF report and nine days before Tania's Estimates appearance, Deloitte Global announced FY2025 revenue of US$70.5 billion and a headcount above 470,000, alongside a commitment of more than US$3 billion to generative AI through FY2030. Its Global CEO, Joe Ucuzoglu, said the firm was “uniquely positioned to help businesses, governments, and markets capture the full potential of AI.” Deloitte sat on the MAS Panel alongside KPMG, EY, PwC, and the major strategy firms; the panel had already survived one scandal, PwC's 2023 misuse of confidential Treasury information, that had hardened Senate Estimates' patience for procurement irregularities.

Deloitte's public AI positioning had two threads: a global Trustworthy AI framework promising safeguards, audit trails, and governance; and a deepening alliance with Anthropic, whose Claude model was marketed on a Constitutional AI design. By the second half of 2025 the two firms were finalising an announcement, already negotiated, slated for 6 October, that would put Claude in front of Deloitte's entire global workforce and certify 15,000 professionals on it. Tania did not yet know the details; the calendar coincidence with her own Estimates appearance three days later was, at this point, no one's plan.

4. The engagement

The December 2024 Order for Services specified two deliverables, a Statement of Program Assurance and a Final Report with improvement options, for A$440,000, with the Final Report due 4 July 2025. Three variations followed: the first added personnel, the second amended milestones, and the third extended the contract end date by a week.

The engagement ran in four phases: document review and interviews, a technical workstream reviewing the TCF's code base against legislative and policy requirements, then synthesis and reporting. For the technical workstream, Deloitte proposed, and DEWR approved after internal assessment, a generative AI tool chain based on Azure OpenAI's GPT-4o, licensed by DEWR and hosted on DEWR's own Azure tenancy, for one purpose only: assisting the technical analysis of code.

Deloitte delivered the Final Report on 4 July 2025: 234 pages, 141 citations, eight key themes, 46 recommendations. It tested 368 of the system's more than 370 business rules and confirmed three defects that had produced inappropriate outcomes for 1,326 participants. Its conclusion: the cumulative effect of the issues found precluded assurance on the Framework's functionality, its legislative alignment, or its governance.

What DEWR did not know was that, between finishing the assurance work and issuing the deliverables, Deloitte's team had used two other generative AI tools, MyAssist (Deloitte's own platform) and ChatGPT, for publishing-related tasks: further summarising a prepared account of a Federal Court judgment, and completing and formatting citations against a style guide. Deloitte's letter later said the outputs had been reviewed before use, and that, in the firm's own words, "while this review corrected some errors, not all were identified as part of this process." DEWR had approved a single tool chain, for one purpose, on its own infrastructure; Deloitte had used two further tools, for different purposes, outside DEWR’s approved environment, and had not told its client (Exhibit C compares the two deployments).

5. The discovery

In early August 2025, Dr Christopher Rudge of the University of Sydney, reading the Final Report for his own research, found a footnote citing a book by his colleague Professor Lisa Burton Crawford. She had not written it. He found a second phantom Burton Crawford book, a nonexistent article credited to a real scholar in a real journal, and a citation attributed to an author whose actual article had a different title and argument than what the report claimed. He contacted the Australian Financial Review, whose journalist sent four citations to DEWR for verification.

The query reached Tania at 1:24 p.m. on Friday, 22 August. She emailed Deloitte at 1:41 p.m., attaching the four citations, copying her senior officials, and asking for an urgent answer. By evening, Deloitte's team called the errors a transcription issue that did not affect the report's substance. On 25 August, Deloitte phoned the Department of Finance directly; DEWR was not on the call. DEWR began its own line-by-line citation review on 29 August. It found that more than 30 of the 141 citations were problematic and 20 were probable hallucinations: invented sources, or real sources garbled beyond recognition.

The most consequential hallucination concerned Amato v Commonwealth, the 2019 Federal Court consent order that anchored the report's argument that automated decisions need probative material behind them. The report cited it as a 2021 immigration judgment with the wrong case number and wrong paragraph references, and attributed to the judge a quotation that appears nowhere in the actual order (Exhibit B catalogues this and the other nineteen hallucinated citations).

6. The substance underneath the citations

By 29 August, DEWR's internal question had shifted: if the citations were unreliable, what about the analysis they supported? Most of the report's 234 pages, on legislative traceability, IT-system defects, governance, and compliance-model design, rested on document review, interviews, and code review; the disputed citations supported only a narrow slice of the analysis. Two load-bearing findings, however, depended on the corrupted legal analysis: that the TCF's path from statute to code to participant outcome was not reliably traceable, and that this untraceability had already harmed 1,326 people. Tania could not know, without a line-by-line review her department was still conducting, whether the corrupted footnotes compromised only the report’s form or reached its substance as well, a question sharpened by the Commonwealth Ombudsman's finding, two weeks earlier, that more than 1,000 recipients had had payments unlawfully cancelled by the very system Deloitte had been asked to assure.

Five options, not mutually exclusive, were on the table by 5 September.

7. The decisions on the table

Option A, narrow contractual remedy: demand return of the A$97,587.11 final instalment, on the ground that the report failed to meet professional quality standards. The remedy is procedurally simple and defensible on the record, and it can be settled quickly. At under a quarter of the contract value, however, it would likely be judged too small by public commentary.

Option B, full refund: demand the entire A$440,000, arguing the citation failure had infected the report's substance. Politically resonant; Senator Pocock would later tell the press that Deloitte's "bungled report isn't just artificial incompetence, it's wilful negligence," and call it "expensive corner-cutting disguised as consultancy." The position was harder to sustain contractually, since Deloitte had committed in writing that its substantive findings were unaffected, and any push for a full refund would likely mean months of dispute or litigation.

Option C, Supplier Code of Conduct escalation: the Code's integrity clause, administered by the Department of Finance through a graduated process of warning, remediation plan, suspension, and termination, could see Deloitte suspended from the MAS Panel for up to five years, a signal Tania's Minister privately favoured. The path would run for months and consume political capital, and it would deprive DEWR, and every other federal agency, of advisory capacity the public service did not maintain in-house.

Option D, a hybrid of narrow remedy now and procurement reform for the future: close the financial loop with Option A while proposing a new doctrine for AI use in future engagements, requiring consultants to declare every AI tool and its purpose upfront, requiring the client department to consent in writing, and designing quality assurance

specifically to catch AI-native errors that differ from the errors junior consultants typically make. The doctrine existed only in internal conversations so far; its principal risk was pre-emption, since the Department of Finance was rumoured to be drafting its own cross-government AI-procurement framework.

Option E, withhold publication of the corrected report: DEWR could decline to publish on the ground that confidence in Deloitte's process was insufficient. The cost: withholding from Parliament and the public findings that, citation problems aside, said DEWR could not assure the lawfulness of its own welfare-conditionality system, a decision unlikely to survive scrutiny once it became known.

The options combined in some ways and not others: A paired naturally with D, and C with B; E combined with A only if Tania could explain why one part of the contract was severable and the other was not. Whatever Tania chose, she would have to explain it twice: once to Deloitte on Tuesday, once to a Senate committee in October.

Sunday, 7 September 2025, found Tania at her desk. The corrected report had not yet arrived; it would come by email on 26 September, with a letter from Deloitte's Chief Risk Officer. The Anthropic announcement had not been made, and Estimates had not convened. Five options remained, and one letter was due Tuesday morning.

Exhibit A: Decision Timeline

Date	Event
26 Nov 2024	Tania Rishniw approves the commissioning of an independent assurance review of the Targeted Compliance Framework.
23 Dec 2024	Order for Services issued under the MAS Panel Head Agreement. Contract value A$440,000.
Feb–Jun 2025	Three Order Variations: personnel added (17 Feb); milestones for stages 3–4 amended; contract end date extended by a week (27 Jun).
18 Jun 2025	Statement of Assurance issued by Deloitte.
4 Jul 2025	Final Report (234 pp, 141 citations) delivered by Deloitte.
Early Aug 2025	Dr Christopher Rudge identifies hallucinated Burton Crawford citations.
22 Aug 2025, 1:24 pm	Australian Financial Review journalist queries four citations with DEWR.
22 Aug 2025, 1:41 pm	Tania Rishniw emails Deloitte requesting an urgent response by close of business.
22 Aug 2025 (late)	Deloitte responds, describing the issue as a transcription error.
25 Aug 2025	Deloitte initiates a phone call to the Department of Finance; DEWR is not on the call.
27 Aug 2025	DEWR contacts Deloitte again.
29 Aug 2025	DEWR commences a line-by-line internal review of the citations.
2 Sep 2025	Deloitte's Lead Partner writes a confidential letter to the Department of Finance.
5 Sep 2025	DEWR–Deloitte meeting; Tania first sees the 2 September letter, forwarded by Finance.
5–7 Sep 2025	**Decision moment for Tania Rishniw.**
9 Sep 2025	Tania writes to Deloitte demanding the A$97,587.11 final-instalment refund and correction of the 2 September letter.

26 Sep 2025	Deloitte delivers the corrected Final Report and revised Statement of Assurance, with methodology disclosure.
30 Sep 2025	Deloitte Global FY2025 announcement: US$70.5 bn revenue, 470,000 staff, US$3 bn GenAI investment.
3 Oct 2025	DEWR publishes the corrected report.
6 Oct 2025	Anthropic and Deloitte announce the expanded global alliance (Claude for 470,000 staff).
8 Oct 2025	Senate Finance Estimates question the Department of Finance.
9 Oct 2025	Senate Education and Employment Estimates: Rishniw and Secretary James testify; James articulates the transparency / agreement / AI-aware QA doctrine.
13 Oct 2025	Digital Transformation Agency releases mandatory guidance on government use of public generative AI tools.

Sources: FOI 25-26/084; DEWR FOI LEX 1828; Hansard, 8–9 October 2025; Deloitte and Anthropic press releases.

Exhibit B: The Hallucinated Citations

Of the Final Report's 141 citations, DEWR's internal review and a retrospective check by GPTZero identified more than 30 problematic citations; 20 were probable hallucinations. Three patterns recurred: a phantom monograph attributed to a real scholar (Professor Lisa Burton Crawford was credited with two nonexistent books; her real book has a different title); a real article with fabricated metadata (a real author and real journal attached to a nonexistent piece, or a real piece cited under the wrong title to support a position its actual author argued against); and a composite person (a named judge who was a blend of two real Australians).

The most consequential case was Amato v Commonwealth, the 2019 Federal Court consent order establishing that automated decisions require probative material. The three versions can be read side by side.

What the judgment actually said (Davies J, VID611/2019, 27 November 2019): “there was no material before the decision-maker capable of supporting the conclusion that a debt had arisen,” and “the conclusion that a debt had arisen was therefore irrational, in the requisite legal sense.”

What the original report said it said (4 July 2025, footnote 57): it cited a 2021 immigration case (Amato v Commonwealth of Australia [2021] FCA 1019, [47]–[50]) with the wrong year, case number, and paragraphs, and put in Davies J’s mouth language he never used, to the effect that “the burden falls on the decision-maker to be satisfied with evidence that the debt is owed” and that “the statutory rights of a person cannot be lawfully reduced on the basis of an unsubstantiated assumption.”

What the corrected report says (26 September 2025): it restores the correct citation and reproduces Davies J’s actual paragraph 9 verbatim.

Sources: GPTZero Inc., company blog, 26 November 2025; Federal Court of Australia, VID611/2019; original and revised versions of the Final Report.

Exhibit C: Deloitte's Generative AI Architecture, as Disclosed

Agreed deployment, phase 3: a ChatGPT-based tool chain on Azure OpenAI GPT-4o, licensed and hosted by DEWR, for the sole purpose of code-base analysis, disclosed at proposal stage with written consent.

Undisclosed deployment, phase 4: MyAssist, Deloitte's private-cloud platform built on GPT-3.5 Turbo and GPT-4o, and ChatGPT, accessed through its public interface, used to summarise a judgment excerpt and to format citations, with no disclosure to DEWR before publication.

Absent from either deployment: retrieval-augmented generation against a verified bibliography, which would have substantially reduced the risk of fabricated citations and enabled systematic verification, particularly if citation generation had been constrained to records in the verified index; an automated citation-verification subroutine checked against bibliographic databases; human-in-the-loop review at the citation level; a persistent, attributable audit trail of prompts; and an explicit scope-segmentation clause prohibiting AI use beyond what the client had approved.

Sources: FOI 25-26/084, Document 1; Final Report (revised), Section 6.3 and Appendix C.

Exhibit D: The Doctrine on the Record (Senate Estimates, 9 October 2025)

One month after Tania's letter, the Secretary of DEWR, Natalie James, set out the department's new posture before the Senate. "You're quite right, Senator; we should not be receiving work that has glaring errors in footnotes and sources. My people should not be double-checking a third-party provider's footnotes," she testified. Quality assurance, she added, "needs to be framed to pick up what AI might get wrong, which might be different to what your junior staff or graduates might get wrong."

Sources: Hansard, Senate Education and Employment Legislation Committee, 9 October 2025.

Teaching Note

Confidential instructor's note. This teaching note is intended for instructor use, and carries the analytical apparatus that the case deliberately withholds from students.

Synopsis

When Assurance Fails follows Tania Rishniw, Deputy Secretary of Australia's Department of Employment and Workplace Relations, in the first week of September 2025, as she confronts a four-part decision she must resolve by Tuesday morning. Nine weeks earlier, Deloitte had delivered a 234-page, A$440,000 assurance review of an automated welfare-penalty system; the report contained twenty hallucinated citations, including a fabricated judicial quotation anchoring its central legal argument. By early September, Deloitte had admitted, to a different government department, in a letter Tania obtained only by inter-agency forwarding, that undisclosed generative AI tools had produced two of the report's publishing tasks. Tania must decide how much money to recover, whether to escalate under the Commonwealth Supplier Code of Conduct, whether to publish the corrected report, and whether to demand correction

of Deloitte's letter: four decisions that the case treats as a single decision about DEWR's institutional posture.

The case is well suited to MBA and executive-education sessions on AI governance, public-sector procurement, professional-services ethics, and the introduction of generative AI into regulated work. It is a public-record case, drawn from Hansards, FOI releases, a Federal Court judgment, and attributable press releases, with no interviews conducted.

Learning objectives

By the end of the discussion, students should be able to: diagnose failure modes in generative-AI-augmented professional deliverables, distinguishing disclosed from undisclosed use and process failures from substantive ones; evaluate GenAI architectures, from plain LLM use to retrieval-augmented generation to human-in-the-loop review, against the assurance properties regulated-sector buyers require; design contractual safeguards for AI use in third-party deliverables; apply accountability frameworks, including algorithmic auditing, Constitutional AI, the OECD AI Principles, and Deloitte's own Trustworthy AI framework, as lenses on procurement-side governance; and weigh contractual remedy, regulatory escalation, and systemic reform as competing institutional responses.

Suggested courses and positioning

The case fits electives in AI governance, best placed after students have covered basic GenAI architecture but before regulatory-design sessions; core or elective courses in public-sector management and procurement; and technology-ethics or business-and-society courses, where it grounds the ethical question in an operational decision the protagonist must act on by a deadline. Suggested pre-readings include Raji and Buolamwini (2019) on algorithmic auditing, Lewis et al. (2020) on retrieval-augmented generation, the OECD AI Principles, and Deloitte's published Trustworthy AI framework. The case runs well in 75 to 90 minutes; below 60 minutes it loses pedagogical density.

Theoretical linkages

Four frameworks anchor the discussion. Algorithmic accountability (Raji and Buolamwini, 2019) frames what Tania does when she traces the firm's AI use through its output and ties remedy to demonstrable harm, useful for the diagnostic stage. Retrieval-augmented generation, or RAG (Lewis et al., 2020), is the architectural concept that organises the redesign exercise: citations grounded in a verified retrieval index substantially reduce the risk of referencing sources that are not in the index, particularly when generation is constrained to that index, the safeguard Deloitte's deployment lacked. Constitutional AI (Bai et al., 2022), Anthropic's alignment-by-training approach for the Claude model Deloitte was simultaneously rolling out to 470,000 staff, offers a counter-example to the deployment chosen for this engagement. The OECD AI Principles (2019, revised 2024) supply the regulatory-design vocabulary, transparency, robustness, accountability, against which the Digital Transformation Agency's October 2025 guidance was later harmonised. Deloitte's own Trustworthy AI framework, with six pillars (fair, robust, transparent, accountable, safe, privacy-respecting), is the most uncomfortable lens, since applying

the firm's published doctrine to its own conduct surfaces a governance gap inside the firm itself.

Question 1: Diagnostic

Question 1, diagnostic: what is Tania really deciding by Tuesday morning? The four discrete decisions (money, escalation, publication, correction) resolve into one question about DEWR's institutional posture. A common error is treating them independently; redirect by asking whether a given combination tells a coherent story. A second error treats this as a question about Deloitte's conduct; redirect to Tania's remit, the integrity of the system she runs.

Question 2: Chronology and the duty to disclose

Question 2, chronology and the duty to disclose: walking the timeline, when should each party have known and disclosed? The instructor should draw two lines, what Deloitte's own published AI policy required at each pivot and what Deloitte actually did, and show that the gap widened as political stakes rose. A common error fixates on the citation errors alone; redirect toward the more serious failure, describing the same problem differently to two arms of government. A second error treats the lapse as partner-specific; surfacing schedule pressure, with phase four compressed and citation work the lowest-status task, makes the failure look predictable: a structural outcome of compressed timelines, repeated across engagement teams under similar pressure.

Question 3: Significant Event

Question 3, Significant Event: do the citation errors meet the contract's materiality test? Strongest answers press on whose materiality counts: for a deliverable whose entire purpose is reliance, and that will be tabled before a Senate committee, public-sector materiality includes reputational and accountability dimensions the firm's letter elides. A common error accepts the firm's framing that materiality turns only on whether substantive findings changed.

Question 4: Architectural redesign

Question 4, architectural redesign: what would retrieval-augmented generation, human-in-the-loop review, citation-verification subroutines, and scope segmentation each have cost, and which of them would have prevented the failure rather than only exposing it? The instructor should surface the productivity-versus-assurance trade-off, then push past it: no architecture helps if the firm chooses not to disclose. A common error redesigns for productivity while ignoring the relational disclosure failure.

Question 5: Whose mandate

Question 5, whose mandate: comparing the five options, which optimises for Tania's narrow mandate, and which for the Australian Public Service's broader one? Option D is best read as a strategy for converting a localised remedy into precedent, with real risks of overreach and pre-emption. A vote is useful here; the class typically splits. A common error favours Option B, full refund, as maximally accountable without weighing the cost of months of dispute; a second favours Option E, withholding

publication, without weighing the political cost of suppressing findings Parliament needs.

Question 6: Trustworthy AI™ applied to its author

Question 6, Trustworthy AI applied to its author: walking Deloitte's own six pillars against its conduct on this engagement shows that the firm's published doctrine, applied to itself, would have required exactly the conduct that did not occur. The payoff is that corporate AI-ethics frameworks fail from a lack of operationalisation: sound doctrine remains unconnected to the procedures, training, and incentives that bind engagement teams in real time. If students stall, ask them to identify, by role, the specific person who should have stopped the engagement at each pivot.

Question 7: The regulatory tide

Question 7, the regulatory tide: did Tania's letter shape the transparency, agreement, and AI-aware quality-assurance doctrine that spread across federal agencies within five weeks, or would that doctrine have arrived regardless, carried by Robodebt, the Ombudsman's finding, the PwC scandal, and the DTA guidance already in drafting? Both forces operated. Isolating a single cause misreads the episode: the doctrine emerged from the letter and the surrounding pressures together.

Board plan and suggested timing

A four-board, 90-minute layout works well. Board 1 tracks the five options and their first- and second-order consequences, built during Questions 1 and 2 and revisited at Question 5. Board 2 compares the AI architecture as deployed against a fully governed alternative, with productivity cost and assurance gain for each, built during Questions 3 and 4. Board 3 maps the disclosure-and-QA doctrine against supplier and buyer obligations, built during Question 5. Board 4 closes with a single generalising sentence the class has earned, echoing the operationalisation gap the discussion surfaced in Question 6. For 75-minute sessions, compress the architecture and decision blocks; below 60 minutes the case loses density.

Epilogue

On 9 September 2025, Tania wrote to Deloitte demanding the A$97,587.11 instalment, correction of the 2 September letter, and confirmation of the firm's remediation process, without invoking the Supplier Code of Conduct. Deloitte agreed, and on 26 September delivered a corrected Final Report and revised Statement of Assurance under its Chief Risk Officer. DEWR published the corrected report on 3 October, without a press release. Three days later Anthropic and Deloitte announced Claude's rollout across Deloitte's global workforce. On 8 and 9 October, Senate Estimates questioned both Finance and DEWR; Secretary Natalie James articulated the transparency, agreement, and AI-aware quality-assurance doctrine on the record, and Minister Murray Watt agreed that a public apology from Deloitte would be appropriate. On 13 October the Digital Transformation Agency issued mandatory generative-AI guidance across government. By December, the doctrine Tania's letter had helped provoke was being written into standard federal procurement templates. Whether her letter provoked the regulatory shift that followed, or coincided with one already under way, is left open by design.

References

Anthropic. (2025, October 6). Deloitte will make Claude available to 470,000 people across its global network [Press release]. https://www.anthropic.com/news/deloitte-anthropic-partnership

Australian Digital Transformation Agency. (2025, October 13). DTA releases new guidance: Australian Government use of public generative AI tools [Media release]. https://www.dta.gov.au/media-releases/dta-releases-new-guidance-australian-government-use-public-generative-ai-tools

Bai, Y., Kadavath, S., Kundu, S., Askell, A., Kernion, J., Jones, A., Chen, A., Goldie, A., Mirhoseini, A., McKinnon, C., Chen, C., Olsson, C., Olah, C., Hernandez, D., Drain, D., Ganguli, D., Li, D., Tran-Johnson, E., Perez, E., and Kaplan, J. (2022). Constitutional AI: Harmlessness from AI feedback [Technical report]. Anthropic. https://arxiv.org/abs/2212.08073

Deloitte Global. (2025, September 30). Deloitte reports FY2025 revenue [Press release]. https://www.deloitte.com/global/en/about/press-room/global-revenue-announcement.html

Deloitte Touche Tohmatsu. (2025a). Targeted Compliance Framework, Independent Review Final Report (4 July 2025). Department of Employment and Workplace Relations. https://www.dewr.gov.au/assuring-integrity-targeted-compliance-framework/resources/targeted-compliance-framework-assurance-review-final-report

Deloitte Touche Tohmatsu. (2025b). Targeted Compliance Framework, Independent Review Final Report, revised (26 September 2025). Department of Employment and Workplace Relations. https://www.dewr.gov.au/assuring-integrity-targeted-compliance-framework/resources/targeted-compliance-framework-assurance-review-final-report

Department of Employment and Workplace Relations. (2025). Freedom of Information disclosure log, FOI LEX 1828. https://www.dewr.gov.au/about-department/corporate-reporting/freedom-information-foi/foi-disclosure-log

Department of Finance. (2025). Freedom of Information disclosure log, FOI 25-26/084, Document 1. https://www.finance.gov.au/about-us/freedom-information/disclosure-log/foi-foi-25-26/084

Federal Court of Australia. (2019). Amato v Commonwealth of Australia (Consent orders, VID611/2019, 27 November 2019). https://comcourts.gov.au/file/Federal/P/VID611/2019/3859485/event/30114114/document/1513665

GPTZero Inc. (2025, November 26). The Deloitte citation situation: $98K controversy explained [Blog post]. https://gptzero.me/news/deloitte-australia-citation-check/

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Kuttler, H., Lewis, M., Yih, W., Rocktaschel, T., Riedel, S., and Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. Advances in Neural Information Processing Systems, 33, 9459-9474. https://arxiv.org/abs/2005.11401

Organisation for Economic Co-operation and Development. (2024). OECD AI Principles (rev. ed.; orig. pub. 2019). https://oecd.ai/en/ai-principles

Parliament of Australia. (2025, October 8). Senate Finance and Public Administration Legislation Committee, Estimates Hansard. https://parlinfo.aph.gov.au/parlInfo/search/summary/summary.w3p

Parliament of Australia. (2025, October 9). Senate Education and Employment Legislation Committee, Estimates Hansard. https://parlinfo.aph.gov.au/parlInfo/search/summary/summary.w3p

Pocock, B., and Allman-Payne, P. (2025, October 9). Greens slam Deloitte's unethical behaviour and measly refund over AI report bungle [Media release]. Australian Greens. https://greens.org.au/news/media-release/greens-slam-deloittes-unethical-behaviour-and-measly-refund-over-ai-report

Raji, I. D., and Buolamwini, J. (2019). Actionable auditing: Investigating the impact of publicly naming biased performance results of commercial AI products. Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society, 429-435. https://dl.acm.org/doi/10.1145/3306618.3314244

Royal Commission into the Robodebt Scheme. (2023). Final report. https://robodebt.royalcommission.gov.au/publications/report