AI Disclosure in Professional Practice

PKG Safety Innovation

PKG-POS-001 | Position Statement | Version 1.0 | 1 January 2026

Preamble

This document sets out PKG Safety Innovation's position on disclosing the use of AI assistants, large language models and generative AI tools in our professional practice.

The landscape is evolving. Norms around AI disclosure are not yet settled, and our own thinking will continue to develop. What follows represents our current position; honest, considered and open to revision as the field matures.

This is not our AI policy. The operational details including approved tools, data handling, security practices, consent processes; live in our Responsible Innovation policy (PKG-POL-002). This document addresses a narrower question: When and how should we disclose that AI was used in our work?

How This Connects to How We Work

PKG operates on a philosophy we call our Safety Innovation OS; our purpose, values, and practices. Two of our values are directly relevant here:

Honesty: We tell the truth, even when it's hard. We tell clients what they need to hear, not what they want to hear. We admit when we don't know. This position statement exists because honesty demands clarity about how we work.

Rigour: We do it the right way, not the easy way. We don't cut corners, even under pressure. The work has to be sound; researched properly, tested and built to hold up. AI makes some things faster. Rigour ensures faster doesn't mean worse.

Lessons from Industry

Recent incidents underscore the importance of AI disclosure in safety professional practice. In 2025, Deloitte delivered AI-assisted reports to both the Australian and Canadian governments containing fabricated citations[1], non-existent journal articles, and in one case, a fictitious quote attributed to a Federal Court judgment. These errors resulted in partial refunds, reputational damage, and regulatory scrutiny.

These weren't failures of AI technology. They were failures of professional oversight, verification and quality control. AI was deployed without adequate human review, fact-checking, or the governance processes that professional services demand.

We cite these cases to emphasise that AI-assisted work requires rigorous quality assurance. As Australian legal firm Spruson & Ferguson noted: "savings made by adoption of AI have to be balanced against the added cost of double-checking the quality of the work."[2]

Our Core Position

When clients engage PKG Safety Innovation, they are purchasing expertise, judgment, and the ability to synthesise complex information into actionable insight. They are not purchasing the mechanical act of typing or the labour of first-draft composition. The value we provide is rooted in decades of combined experience across safety, health, environment and risk management, human factors, digital transformation, and technology deployment.

AI tools, including large language models, are instruments that enhance our capacity to deliver that expertise more comprehensively and efficiently. Just as we would not disclose the use of spellcheck, reference management software, or a research assistant, we do not believe that every use of AI-assisted tooling requires explicit disclosure at the point of output.

> *The relevant question is not 'was AI used?' but rather 'does the work reflect genuine expertise, rigorous quality control, and professional accountability?'*

Unlike the cases cited above, we never use AI to blindly generate content or conduct research without domain expertise informing and validating the output. AI synthesis is directed by professional judgment and subject to verification against primary sources.

Regulatory & Standards Alignment

PKG Safety Innovation operates from Australia with a global client base. We follow Australian guidance as our baseline. Where international standards offer greater rigour or clarity, we adopt those frameworks.

Our AI practices align with:

- **Australian Voluntary AI Safety Standard** — particularly guardrails on transparency, human oversight, and accountability
- **NIST AI Risk Management Framework (AI RMF 1.0)** — the US framework covering governance, risk mapping, measurement, and management
- **ISO/IEC 42001:2023** — the international standard for AI management systems and the associated **ISO/IEC 29843:2023** Guidance on AI Risk Management.

As practitioners within the health and safety profession, we also uphold the Australian Institute of Health and Safety (AIHS) Code of Ethics. Integrity requires that we do not misrepresent the nature or provenance of our work. Honesty requires that we are transparent with clients about our methods. Accountability requires that we take full professional responsibility for all deliverables regardless of the tools used in their creation.

Our Disclosure Framework

We apply a tiered approach to disclosure based on context, the nature of the work, and the reasonable expectations of our audience.

Client Engagements & Consulting Deliverables

For all client engagements, we:

- **Obtain consent for the use of AI assistants** as part of our engagement terms
- **Disclose in proposals** that large language models will be used in research, synthesis, and reporting
- **Maintain full professional accountability** for all deliverables regardless of the tools used in their creation
- **Apply rigorous verification and fact-checking** to all AI-assisted outputs

This upfront transparency satisfies disclosure obligations without cluttering individual reports with repetitive disclaimers. Our clients engage us precisely because we integrate AI capability effectively; it is a feature of our service.

Visual Content

When we generate images, diagrams, or visual frameworks using AI tools, we disclose this. Visual content carries different expectations around authenticity and provenance than text-based analysis. Where visual content has been substantially created by AI, this is noted.

Public Content & Thought Leadership

For thought leadership content including LinkedIn posts, blog articles and podcast-related materials we generally do not append explicit AI disclaimers. This content represents our professional perspective, informed by our expertise and refined through whatever tools assist clear expression. The ideas, arguments, and positions are authentically ours; the drafting process is incidental.

We reject both extremes of the current discourse: the defensive '100% human-generated' badge (which implies AI assistance is somehow shameful) and the overwrought disclaimer that treats AI collaboration as confession. Neither reflects the reality of modern professional practice.

Disclosure Rationale

What Is Actually Being Claimed?

The purpose of disclosure is to prevent deception; to ensure that claims being made are accurate. When we publish analysis of safety transformation discovery, data insights or critical risk management practices, we are claiming expertise in these domains. We are not claiming to have personally typed every word. The former matters; the latter does not.

AI Detection Software

Current AI detection tools are unreliable. They pattern-match on stylistic features, meaning that clear, well-structured writing frequently flags as 'AI-generated' regardless of its actual provenance. We have experienced this directly: content written entirely without AI assistance has been scored as highly AI-generated. A disclosure regime built on detection technology that cannot accurately distinguish human from AI writing is fundamentally flawed.

Technology Continuum

Consider the spectrum of writing assistance: spellcheck and grammar correction, thesaurus and style suggestions, research databases and literature review tools, human editors and writing coaches, AI-assisted drafting with expert oversight, and fully automated generation without human review.

The first four are used universally without disclosure. The last, exemplified by the Deloitte failures, is clearly different in kind and represents practice we explicitly reject. AI-assisted work by domain experts sits between these poles; and, in our view, much closer to the 'enhanced tools' end when the human is providing expertise, direction, and quality control.

Where Disclosure is Required

We believe explicit disclosure is appropriate when there is an implicit claim that AI assistance would undermine:

- Personal experience or testimony (memoirs, first-person accounts)
- Academic submissions with explicit institutional rules
- Journalism with source attribution standards
- Legal or regulatory filings with specified requirements
- Contexts where clients or audiences have explicitly requested disclosure

Professional consulting deliverables, where clients have been informed of our AI-integrated approach and are paying for outcome quality, do not fall into these categories.

Quality Assurance Commitments

Learning from industry failures, we commit to the following for all AI-assisted work:

1. **Source verification:** All citations, references, and attributed statements are verified against primary sources before inclusion in deliverables.
2. **Domain expert review:** AI-generated content is reviewed by qualified practitioners with relevant subject matter expertise.
3. **Factual accuracy checks:** Statistical claims, regulatory references, and technical assertions are independently verified.
4. **Professional accountability:** All deliverables are signed off by accountable practitioners who take full responsibility for content accuracy.

We also provide ongoing education and training for our team as part of our Innovation OS.

Looking Forward

We anticipate that within several years, the current anxiety around AI disclosure will appear as dated as someone declaring their document was 'handwritten, no word processor used.' The question will shift from 'was AI involved?' to 'was the work good?'

We position ourselves ahead of that curve while remaining respectful of current transitional concerns and regulatory expectations. As the Australian Government moves toward mandatory guardrails for high-risk AI applications, and as international standards continue to develop, we will adapt our practices accordingly. This position statement will be reviewed and updated as norms evolve, technology capabilities change, and regulatory requirements crystallise. For now, this is where we stand.

Our Commitment

Regardless of the tools we employ, our commitment remains constant: to deliver work of the highest professional standard, grounded in genuine expertise, subject to rigorous quality control, and backed by full professional accountability.

We welcome dialogue with clients, colleagues, regulators, and industry bodies on these questions.

PKG Safety Innovation
A Pocketknife Group Practice
1 January 2026

Reference Framework

Australian Guidance

- Australian Government Voluntary AI Safety Standard (September 2024)
- Australian Government Guidance for AI Adoption (October 2025)
- Proposed Mandatory Guardrails for AI in High-Risk Settings

International Standards

- NIST AI Risk Management Framework (AI RMF 1.0) — January 2023
- ISO/IEC 42001:2023 — AI Management Systems

- GDPR Regulation (EU) 2016/679

Professional Ethics

- Australian Institute of Health and Safety (AIHS) Code of Ethics

Related PKG Documents

- PKG-POL-002 Responsible Innovation — AI use, privacy, data handling, security

References:

1) Paoli. N (2025). *"Deloitte allegedly cited AI-generated research in a million-dollar report for a Canadian provincial government"* Fortune. https://fortune.com/2025/11/25/deloitte-caught-fabricated-ai-generated-research-million-dollar-report-canada-government/
2) Tso, S & Martino, N. (2025). "Australia | Responsible AI use & lessons from Deloitte's "Dodgy Report"". https://www.spruson.com/australia-responsible-ai-use-lessons-from-deloittes-dodgy-report/