Code of Conduct: Responsible Use of AI in Expert Reports on the Mental Functioning of Offenders

Version: 1.0
Initial draft prepared by: Peter Hanley, Psychologist, Melbourne Private Psychology
Date: 18 August 2025
Web: https://github.com/melbpsych/ai

---

1. Introduction

1.1 Purpose

This Code establishes standards for the ethical and responsible use of Artificial Intelligence (AI) in preparing expert psychological and psychiatric reports concerning the mental functioning of offenders. It aims to ensure accuracy, fairness, transparency, and compliance with professional obligations in forensic contexts, while enabling experts to leverage AI to enhance report quality and efficiency.

1.2 Scope

This Code applies to:

- Psychologists and psychiatrists preparing reports for courts, tribunals, parole boards, or other forensic settings.
- Legal practitioners commissioning expert reports.
- Judicial officers evaluating the reliability of expert evidence.

1.3 Application

- The Code is self-regulatory and does not replace any other codes or laws that may be applicable to the expert.
- Experts adhering to this Code may include a compliance declaration in their reports.
- Hosted on GitHub, the Code supports open access, commentary, and version-controlled updates.

---

2. Core Principles

Human Accountability

The expert is solely responsible for all opinions, analyses, and conclusions in the report. AI tools must never replace professional judgment.

Transparency

Any use of AI in the report preparation process must be explicitly declared.

Accuracy and Verification

All AI-generated outputs must be independently verified against primary sources to ensure factual accuracy and professional integrity.

Confidentiality and Privacy

Sensitive or identifying data must only be processed using secure, enterprise-grade AI systems compliant with relevant privacy regulations, including the [Australian Privacy Principles](#).

Fairness and Impartiality

AI tools must not introduce bias, prejudice, or distortions. Experts remain bound by their duty to provide impartial, objective evidence to the court.

---

3. Practical Guidelines

3.1 Permitted Uses of AI

AI may be used to support the following tasks, provided safeguards are followed:

- **Administrative Tasks:** Formatting, grammar correction, spelling checks, and citation management.
- **Transcription:** Converting audio interviews to text using secure, confidential platforms.
- **Literature Review:** Summarising peer-reviewed studies, with all outputs verified against original sources.
- **Workflow Assistance:** Generating report structures, checklists, or prompts to ensure completeness and conformity to professional and legal standards.
- **Quality Assurance:** Checking drafts for accuracy, consistency, coherence, or potential bias.
- **Exploratory Hypotheses & Alternatives:** Producing draft alternatives, possible lines of reasoning, or highlighting overlooked indicators. AI may be asked to brainstorm potential critiques, generate rival hypotheses, or anticipate challenges under cross-examination.
    - These outputs are not opinions and must be treated only as *ideas for consideration*.
    - The expert must critically evaluate such material and adopt it only if it withstands independent scrutiny and aligns with evidence, professional standards, and judgment.

3.2 Prohibited Uses of AI

AI must not be used for:

- Drafting or generating final professional opinions, diagnoses, or conclusions.
- Analysing raw data (e.g., psychometric tests or interview notes) without independent expert interpretation.
- Processing sensitive or identifying information via non-secure or public AI platforms that do not conform to Australian Privacy Standards.
- Presenting AI-generated text as expert opinion without disclosure and verification.

3.3 Required Safeguards

- **Tool Register:** Maintain a record of AI tools used, including tool name, version, date, and purpose.

- **Audit Trail:** Document all prompts and AI outputs relied upon during report preparation.
- **Data Protection:** De-identify all inputs to AI systems where feasible and use only privacy-compliant tools.
- **Independent Verification:** Cross-check all AI-generated outputs (facts, references, or suggestions) against original sources.
- **Bias Review:** Evaluate AI outputs for potential bias, stereotyping, or inappropriate language.
- **Critical Evaluation:** AI-generated hypotheses or alternatives must be considered *tentative only* and validated against professional reasoning and evidence before inclusion in a report.

---

4. Practical Tools and Examples

4.1 Sample AI Prompts

Consistency Check
"Analyse this draft report for inconsistencies between the client's history, psychometric results, and clinical formulation. Provide a list of discrepancies without modifying the text."

Bias Review
"Review this report summary for language that may imply guilt, blame, or moral judgment beyond the scope of professional assessment. Highlight problematic phrases."

Completeness Check
"Compare this draft report against the provided assessment instructions. Identify any missing or incomplete sections without altering the content."

Exploratory Alternatives
"Suggest possible alternative formulations or diagnostic considerations that might differ from the current draft. Do not provide a conclusion; present them as options only."

Cross-Examination Stress Test
"Identify potential weaknesses or areas of vulnerability in this draft report that an opposing counsel might raise during cross-examination."

4.2 Recommended Workflow

1. Conduct interviews and assessments; record notes manually or with secure methods.
2. Use secure transcription AI (if applicable) and verify transcript accuracy.
3. Draft the report manually, incorporating professional analysis and conclusions.
4. Use AI to generate alternative ideas, criticisms, or stress tests.
5. Critically evaluate AI outputs; adopt only those that are evidence-based and align with expert judgment.
6. Apply AI tools for structure, grammar, and consistency checks as needed.
7. Independently verify all AI-suggested changes or outputs.
8. Finalise the report and include the AI use declaration.

---

5. Summary for Inclusion in Reports

AI Code of Conduct Summary

- The expert retains full responsibility for all report content.
- AI may assist with administrative tasks, secure transcription, literature support, quality checks, and generating alternatives or criticisms.
- AI may brainstorm hypotheses or challenges, but these must be treated as tentative ideas and critically evaluated before use.
- AI must not generate or replace professional opinions or analyses.
- All AI use must be transparent, declared, and independently verified.
- Confidential data must be processed only through secure, compliant systems.

---

6. Versioning and Updates

- This Code is hosted on GitHub to facilitate open access, community feedback, and version control.
- Updates will be tracked with version numbers and documented changelogs.
- Experts must cite the specific version of the Code followed in their reports (e.g., Version 1.1).

DISCLOSURE

This report was prepared in compliance with the *Code of Conduct for the Responsible Use of AI in Expert Reports on Mental Functioning of Offenders (Version 1.0)*, available at https://github.com/melbpsych/ai.

AI tools used: [Name and version]
Purpose of use: [e.g., transcription, formatting, literature summarisation, quality checks, exploratory hypotheses]
Data handling: [e.g., de-identified inputs, secure system, training disabled]
Verification: All AI outputs were independently checked against source materials and critically evaluated.
Limitations: [Any relevant notes]

All professional opinions, analyses, and conclusions are my own, independently verified, and free from unverified AI influence.