Ethical Use of AI in Psychological Report Writing: Viewer Notes

About this session

AI can now draft a whole psychological report, and you remain the author of every sentence in it. This session, recorded for NovoPsych in November 2026 for psychologists in Australia and the United States, is about what that authorship requires.

Presenter: Adam B. Lockwood, PhD, NCSP, LP, Associate Professor of School Psychology at Kent State University and owner of Lockwood Educational and Psychological Consulting. Read more at lockwoodconsulting.net/blog.

Use these notes alongside the recording, or on their own as a reference. They follow the session's seven sections and close with the checklist and some questions to discuss.

Disclosures. Dr. Lockwood has served as a paid consultant for multiple tech companies in the AI space. He received an honorarium from NovoPsych for this presentation. He does not endorse any specific AI product.

AI use. Generative AI helped draft the slides and these notes. Every statistic was checked against its published source, and Dr. Lockwood reviewed the content and remains responsible for it.

Not legal advice. This session offers ethical and professional guidance grounded in professional codes and published research. Questions about law belong with your counsel or your professional indemnity insurer.

Key takeaways

The draft can be generated, but the judgment cannot be delegated, and the signature is still yours.

1. **The profession's caution about AI in reports is well founded, and it is compatible with using AI.** Take the concern seriously, then build the review process that answers it.
2. **Unsupported content is the failure mode you are most likely to produce.** There is nothing to fact-check, so only someone who knows the file catches it.
3. **Competence is the hinge.** If you cannot evaluate the output, that is where you stop, not where you verify harder.
4. **Ownership is a checklist, not a rewrite.** A short review at three set moments does more than general vigilance.

Section by section

Framing: an overly confident intern

Technology is neither good nor bad, nor is it neutral (Kranzberg, 1986). A report-drafting tool is not ethical or unethical on its own; the review process around it makes it one or the other.

Think of AI as an overly confident intern. It is fast and produces polished work from a short instruction. It also starts from zero context every time, rarely asks when it is unsure, and sounds surest where it knows least. Its novice errors do not look like novice errors, so the supervision instincts you trained on do not transfer. And now the intern can hand you a finished report.

1. What the 2023 models already showed

In a study by Lockwood et al. (2025), 249 licensed psychologists blindly rated reports written by four psychologists and reports generated by a 2023 general-purpose model from a single prompt. The reports were generated in 2023 and the ratings collected in early 2024.

- Humans averaged 2.5 hours per report; the model averaged 91.5 seconds.
- Readability, jargon, and sentence length showed no significant difference. Writing style and organization favored humans, with small effects.
- **Recommendations favored the model:** 73% rated its recommendations Average or higher, against 55% for the human ones. Summaries favored humans.
- Raters preferred the human report 49% of the time and the AI report 36%.

Treat these numbers as a floor, not a ceiling. Surface polish was already indistinguishable, so it is not a way to judge quality. The section where the model did best, recommendations, carries the most clinical judgment. Build a review process that does not depend on a quality gap, because the gap is the part that moves.

2. Where the field is now

A November 2025 survey of 518 licensed health service psychologists (Lockwood et al., in press) found that 57% had used AI for work in the past six months. Users reported a median of 2 hours saved per week, and 97% said they review or edit AI content before using it.

Psychologists already sort tasks by risk. Among AI users, 40% used it to answer work questions, but only 15% to write reports and 7.5% to interpret test results.

"It is ethical to use AI to..."	Disagree	Agree
Conduct treatment or intervention planning	33%	49%
Write session notes	44%	38%
Write reports	52%	30%
Interpret test scores and client data	53%	29%

A majority of your colleagues do not consider it ethical to use AI to write reports. Use is driven by usefulness rather than trust: 55% agreed AI is helpful, while 20% agreed with "I trust AI." Low trust may be protective, because it keeps scrutiny high.

The survey's authors note that respondents may be more engaged with AI than the full membership.

3. Two ways it fails

Hallucination is a false claim, such as a wrong subtest count or an invented citation. You can catch it by checking a source.

Unsupported content is true in general but not warranted by this client's data, such as a real, evidence-based treatment with no basis in the evaluation. This is the failure mode you are most likely to produce, for four reasons:

1. There is no error to notice; the sentence holds up on its own terms.
2. Only the person who knows the file can catch it.
3. It survives every automated check, because no tool knows what is in your evaluation.
4. It is the content a reader acts on.

Related risks covered in this section:

- **Missing hedges.** A rewrite can turn "may be consistent with" into "indicates." Read for changed meaning, not changed words.
- **Correct is not the same as competent.** If a licensing board asked you to explain a recommendation, being right would not be enough. AI makes it easy to sound competent outside your actual competence.
- **Anchoring.** Once you read a fluent draft, your review narrows to whether it is acceptable rather than what is correct. You cannot notice the recommendation that was never generated.
- **Agreement.** Prompt a model with your hypothesis and it will support your hypothesis. Ask it instead what the data would look like if you were wrong.
- **Bias.** Models have given different treatment recommendations when a client's race was implied or stated (Bouguettaya et al., 2025).
- **Deskilling.** The risk is not using AI once; it is repeatedly outsourcing the same cognitive work (Hoff, 2011). The skill that detects unsupported content is the one you stop practicing.

A four-question test for any AI-assisted paragraph:

1. Can I point to the specific data that supports this sentence?
2. If I deleted it, would anything in my findings change?
3. Could I explain it without the model's help?
4. Did it come from the client's data, or from the model's priors?

4. The premise is shifting

Until recently, AI produced fragments and you composed the report. Increasingly the draft may arrive complete, so you enter the process at review rather than at authorship. For many clinicians, writing is where the clinical thinking happens. Losing that step changes what review has to do.

5. Using it ethically

You already know what a well-supported report requires, and almost everyone already reviews. The failures happen anyway, so the gap is not knowledge or conscientiousness. Gawande (2009) calls this a failure of ineptitude: you know what to do but do not reliably do it, and the remedy is structure.

Surgery and aviation solved this with a designated pause rather than more training. Build three timeouts into your report workflow: **pause** at a set point, **read** the draft against your own notes rather than the reverse, and **reflect** on what you would have written and what is missing. Expect this to take about five minutes per report.

6. A review checklist

The checklist uses a do-confirm format: do the work from your own expertise, then stop and confirm what is most dangerous to have missed. It focuses on killer items, which are high stakes and hard to detect: unsupported content, competence you cannot demonstrate, and the option the model never generated. Ordinary errors such as wrong dates or invented citations need a verification habit rather than a checklist item.

When the AI disagrees with you, do not treat the model as an independent expert vote. Work out which facts or assumptions produced the disagreement, return to the evidence and professional standards, consult a colleague when the stakes justify it, and document your own reasoning. Disagreement is information about the inputs, not a second opinion.

7. What the Ethics Code already says

The APA Ethics Code (2017) anticipated automated interpretation:

- **9.06** covers interpretation "including automated interpretations" and requires you to state significant limitations. An AI draft will not volunteer its own.
- **9.09(c):** psychologists retain responsibility whether they score and interpret tests themselves or use automated services. **9.09(b)** asks you to select such services on evidence of validity, which is a fair question for any vendor.
- **9.10:** explaining results is your responsibility, whoever produced them.
- **9.01(a):** opinions rest on information sufficient to substantiate the findings. This is the standard for unsupported content.
- **9.02 and 9.03:** use techniques in ways the research supports, and obtain informed consent that covers third parties. Whether an AI vendor counts as a third party is an open question.

One caveat: 9.09 was written for rule-based scoring, where the same input gives the same output. Generative models are not that, so treat it as the closest precedent rather than a settled answer.

US and Australian standards compared

All three bodies agree that responsibility stays with you; the Australian Psychological Society (APS) is more specific about consent and records. Ahpra is Australia's registration regulator, closer to a US state licensing board. The APS is a professional association, the closer counterpart to the APA.

Topic	APA (US)	Ahpra (Australia)	APS (Australia)
Responsibility	You retain responsibility for automated interpretation	"Practitioners must apply human judgment to any output of AI."	You remain fully responsible for any outputs generated by technology.

Topic	APA (US)	Ahpra (Australia)	APS (Australia)
	(Standard 9.09(c)).		
Understanding the tool	Select services on evidence of validity (9.09(b)).	Know the tool's intended use, training and testing, and limits.	Apply evidence-based standards; verify outputs and any references cited.
Disclosure	Substantial use needs more discussion than subtle use (APA, 2025).	What you tell clients "will depend on how and when AI is being used."	Identify AI-produced content in the clinical record.
Consent	Consent covers the nature and purpose of the assessment and third parties (9.03).	Get consent before AI uses client data, and ideally note the response in the record.	Written consent, specific to each tool; a blanket notice is not enough. Consent is ongoing.
Client data	Not addressed directly.	Know whether client data trains the model and where it is stored.	Avoid platforms that reuse client data for training.

If you practice in both countries, the Australian standard is the stricter one to meet.

The APS guidelines themselves are available only to APS members. The APS column draws on the APS's published summary of them (11 February 2026). Ahpra's page was last reviewed in August 2024, so check for updates.

The seven confirmations

Seven checks at three moments catch the failures a careful read misses. This is a reasoned proposal, not a validated instrument, so adapt it: a checklist you did not write is one you will not use.

Before you generate

- ☐ I am using an approved environment with an agreement that covers client data, and the client's consent covers this use.
- ☐ I have written my own independent read of the case: a few sentences on what the data mean, before opening the draft.

At review (read the draft against your notes, not the reverse)

- ☐ I can point to the data that supports each sentence. *(Catches unsupported content.)*
- ☐ I could explain and justify this to my licensing board without the model. *(Catches competence gaps.)*
- ☐ I have checked what is missing that I would have included. *(Catches the narrowed option set.)*
- ☐ No sentence claims more certainty than the data support. *(Catches missing hedges.)*

Before signature

- ☐ I am prepared to be named the author.

The independent read is the most important item. It is the only one that can defeat anchoring, because once you have read a fluent draft you cannot un-read it.

What the checklist will not do: it does not require you to rewrite the draft or distrust the tool, and it will not catch bias on its own. It is not a policy and does not replace legal advice.

Disclosure and documentation

One accurate sentence in your report, plus a short note in your own file, covers most of what disclosure and documentation require. In the 2025 survey, 73% of psychologists said AI use should be routinely disclosed, yet median reported disclosure among AI users was 10%, and the most common answer was zero.

A sample methods statement

> *"[Tool name] was used to generate portions of this report. All findings, interpretations, and recommendations were determined by the examiner."*

- Name the tool, what it did, and what remained yours.
- Say what is true. If AI generated candidate recommendations that you chose among, say so; that is a different contribution from editing prose.
- Write your own sentence, keep it, and revisit it when your tools change.

What to record in your own file

Record	What it covers
Tool	Which platform, and under what agreement
Task	What it was used for, and what it was not used for
Data	What information was provided to it
Review	That you reviewed, edited, and verified the output
Authority	That the professional conclusions are yours

The test: if a client asks in three years whether AI touched their report, you should be able to answer.

Consult your organization's policy, and your counsel or professional indemnity insurer, when you develop disclosure language.

Questions to discuss

Use these on your own, in a consultation group, or with a practice team. Expect disagreement; that is the point.

Where would you draw the disclosure line? For each use, decide whether it needs disclosure in the report, consent beforehand, both, or neither.

1. AI fixed grammar and sentence length in your finished narrative.
2. AI rewrote your summary at an eighth-grade reading level for the client.

3. AI generated five candidate recommendations, and you selected three.
4. AI summarized a 40-page outside evaluation that you then read.
5. AI suggested an interpretation you had not considered; you disagreed and did not use it.

Item 3 is the one most people get wrong: choosing among AI-generated options still lets the model define the option set.

For reflection

- If AI drafted your whole report, would you sign it? What would have to be true first?
- Which of your professional skills are you willing to lose to a tool, and which are you not?
- Where in your current workflow is there a designated pause before you sign?
- In which areas could AI make you sound competent beyond your actual competence?
- What would you need to know about a tool before you could describe its limitations in a report?

References

American Psychological Association. (2017). *Ethical principles of psychologists and code of conduct* (2002, amended effective June 1, 2010, and January 1, 2017). https://www.apa.org/ethics/code

American Psychological Association. (2025). *Ethical guidance for AI in the professional practice of health service psychology* (updated July 2025). https://www.apa.org/topics/artificial-intelligence-machine-learning/ethical-guidance-ai-professional-practice

Australian Health Practitioner Regulation Agency. (2024). *Meeting your professional obligations when using Artificial Intelligence in healthcare.* https://www.ahpra.gov.au/Resources/Artificial-Intelligence-in-healthcare

Australian Psychological Society. (2026, February 11). *Using AI in psychological practice: How the new APS practice guidelines help psychologists.* https://psychology.org.au/insights/use-ai-in-practice-new-aps-practice-guidelines

Banerjee, S., Agarwal, A., & Singla, S. (2024). *LLMs will always hallucinate, and we need to live with this.* arXiv:2409.05746. https://arxiv.org/abs/2409.05746

Bouguettaya, A., Stuart, E. M., & Aboujaoude, E. (2025). Racial bias in AI-mediated psychiatric diagnosis and treatment: A qualitative comparison of four large language models. *npj Digital Medicine, 8*(1), 332. https://doi.org/10.1038/s41746-025-01746-4

Farmer, R. L., Lockwood, A. B., Floyd, R. G., & Sisco, A. D. (2026). AI in psychoeducational assessment: A descriptive study of report generation. *Journal of Applied School Psychology.* https://doi.org/10.1080/15377903.2026.2629395

Gawande, A. (2009). *The checklist manifesto: How to get things right.* Metropolitan Books.

Hoff, T. (2011). Deskilling and adaptation among primary care physicians using two work innovations. *Health Care Management Review, 36*(4), 338–348. https://doi.org/10.1097/HMR.0b013e31821826a1

Kranzberg, M. (1986). Technology and history: "Kranzberg's laws." *Technology and Culture, 27*(3), 544–560.

Lockwood, A. B., Farmer, R. L., & McClintock, J. B. (in press). *Helpful but not trusted: Artificial intelligence adoption and ethical concerns among health service psychologists. Professional Psychology: Research and Practice.*. https://osf.io/preprints/psyarxiv/gf9sq_v1

Lockwood, A. B., Farmer, R. L., Shergill, G., Benson, N. F., & Gilbert, K. (2025). Human vs. machine: Comparing AI-generated and human-written psychological reports. *Journal of Psychoeducational Assessment, 43*(6), 559–573. https://doi.org/10.1177/07342829251346623

National Academy of Medicine. (2025). *An artificial intelligence code of conduct for health and medicine.*

Xu, Z., Jain, S., & Kankanhalli, M. (2024). *Hallucination is inevitable: An innate limitation of large language models.* arXiv:2401.11817. https://arxiv.org/abs/2401.11817