Counterbench.AI — One-page kit

Defensible AI Document Review Protocol (2026) + Tool Shortlist

https://counterbench.ai/guides/defensible-ai-doc-review-protocol-2026 · Last updated: 2026-03-08

QUICK ANSWER

AI-assisted review can be defensible in 2026 if you use a written protocol, require cite-backs to the underlying text, keep batch/decision logs, and run bucketed QA sampling with humans owning privilege and responsiveness decisions.

BENCH-TESTED CHECKLIST

- Define what AI is allowed to do (triage, extraction, draft notes) and what it is not (final privilege calls, auto-production).
- Write the coding guide (responsive/non-responsive, privilege basis categories, issue tags).
- Set data boundaries: approved systems, prohibited systems, and who can export.
- Adopt the cite-back rule: if it can’t point to the text, it’s a draft.
- Stage A: Triage (structured summary + risk flags + escalate Y/N).
- Stage B: Substantive review (humans apply the coding guide; AI assists).
- Stage C: QA sampling (per-bucket sampling; log error types).
- Stage D: Escalation (clear rules; record decisions and changes).

Get templates + the full workflow: https://counterbench.ai/guides/defensible-ai-doc-review-protocol-2026