Counterbench.AI — One-page kit

AI Privilege Log Triage (2026): Failure Modes + Tool Shortlist

https://counterbench.ai/guides/ai-privilege-log-triage-2026 · Last updated: 2026-03-08

QUICK ANSWER

Use AI for privilege log triage only when outputs are cite-backed, email chains are segmented, roles are mapped, and you QA-sample the “non-privileged” and “potential privilege” buckets before final calls.

BENCH-TESTED CHECKLIST

- Require a privilege basis field (attorney-client / work product / common interest / etc.).
- Enforce cite-backs for any privilege rationale used downstream.
- Maintain a role map (names → roles) and keep it updated during review.
- Treat attachments as separate review items; don’t rely on email-only summaries.
- Segment forwarded chains: identify where the privileged portion begins/ends.
- Sample the “non-privileged” bucket—this is where privilege misses tend to live.
- Sample “potential privilege” outputs for over-inclusion patterns.
- Track error types (attachments, chain boundaries, role confusion) and adjust inputs/rules.

Get templates + the full workflow: https://counterbench.ai/guides/ai-privilege-log-triage-2026