AI Guardrails for Tax Teams

Checklist for defensible AI use in tax advisory, compliance, and disputes

1) Scope and permitted use

- **Use AI for**: first-draft writing, outlines, formatting, and summarising **non-confidential** material.
- **Do not use AI for**: final technical conclusions, filing positions, or dispute submissions without human sign-off and verification.
- Treat AI output as draft work product, not authority.

2) Approved tools only

- Maintain an **approved tool list** (enterprise / vetted environments).
- Public or free tools are **not approved** for client work or sensitive matters.
- Access should be role-based; usage should be logged where feasible.

3) Data handling and confidentiality

- **Hard rule**: no client data in public AI tools.
- Apply an Input Sanitisation Standard: remove identifiers; use placeholders (ClientCo, Country A, Transaction X).
- If real client facts are needed: use only approved environments and follow client consent / confidentiality protocol.

4) Citation and authority controls

- **Primary-source rule:** No AI-generated citation or proposition is used unless verified in a primary source.
- Reconcile: (i) authority exists, (ii) proposition matches the authority, (iii) save source reference in the file.
- If not verifiable quickly: remove it and mark as requires verification.

5) Human review and sign-off

- Drafting-only: reviewer checks accuracy and sensitivity.
- Technical analysis: **second-person review** is mandatory (treat like junior work).
- High-stakes items (opinions, disputes, APA/TP positions): lead/partner sign-off required.

6) Workflow controls

- Classify task risk (drafting vs technical vs filing/dispute).
- Sanitize inputs (or stop).
- Generate draft using approved tool.
- Validate facts and authorities.
- Human technical review.
- Finalize and store support trail.

7) Documentation and audit trail

- Store a short AI Use Note in the file: tool used; what AI did; what humans verified; sources checked.
- Keep primary sources (links/PDFs) for any authority relied upon.

8) Training

- Refresh quarterly on hallucinations, outdated knowledge, and automation bias.
- Reinforce input discipline and confidentiality rules.
- Define escalation points: when to stop AI use and route to a technical reviewer.

9) Incident response

- Freeze dissemination; notify engagement lead and risk/compliance.
- Document what was shared and where; remediate deliverable; strengthen the failed control.

10) Client communication boundary

- AI is a productivity tool, not a substitute for professional judgement.
- If asked, confirm that outputs are reviewed by qualified professionals and authorities are validated from primary sources.

Implementation tip: Start with two bright lines - (i) no client data in public AI, and (ii) no unverified AI citation leaves the team.