Corresponding essay: **E18** “When an Agent Acts, Who Acted?”

AGENT IDENTITY CHAIN

The Agent Identity Chain

Bring one consequential action your agent performs. Answer all seven links from credentials and logs.

BRING one agent action you would not want to read about on the front page

1. The seven links, for one action

Answer each link from credentials and logs, not from memory. If you cannot answer it, you do not yet know who acted - and who acted decides who pays.

Link	Your answer for this action	Source of the answer (credential / log / mandate)	Gap?
1. Who authorized it - the human or organizational principal, bound to a credential and a time			
2. What it was allowed to do - scoped mandate: amount, merchant, tool, time window, data classes, irreversibility limits			
3. Which tool acted - the exact integration, API call, browser session, sub-agent or workflow step			
4. What data it touched - inputs, retrieved context, sensitive fields, anything disclosed onward			
5. What decision led to the action - a decision record reconstructable from logs, not only self-report			
6. Was it reversible - refund, undo, recall, delete, escalation route, or an explicit "final" marker			
7. Who is accountable - the named party owning the loss, dispute handling, compliance response and remedy			

2. The two-principal check

User permission and platform authorization are two different things. Ask both authorization questions, not one.

- [ ] We can state what authority the user granted the agent (the delegation question).
- [ ] We can state what authority the receiving platform granted this agent or class of agents (the platform question).
- [ ] The agent is not relying on a borrowed logged-in session treated as complete authorization.
- [ ] A credential travels with the action, and someone on the receiving end can verify it.
- [ ] Accountability is split where the authority is split (e.g. queue policy, credit authority, autonomy level held by different owners).

3. Identity chain by action class

Repeat one row per action class your agent performs. Risk-tier each: a reversible action gets a log and an undo; an irreversible one gets a mandate and a dispute route.

Action class	Risk tier (low / medium / high)	Authoriser and mandate	Receiving platform's authorization	Reversibility and dispute route	Accountable owner

4. The next piece of work

Take one action where the credential answer is "nothing" and the verifier answer is "no one". Write what you will build first.

Companion worksheet to **Essay 18 · When an Agent Acts, Who Acted?**, in the series **Architecting the AI Coworker**. · Dr Peter McCann Strain · Fill this in against one real agent, action class or vendor.

Series Companion + all 22 worksheets: **Release_v12/Series_Companion.pdf**