



THE AI ACCOUNTANT
Your Practice, powered by AI

FREE RESOURCE

The Encoding Loop Starter Kit

Build the loop the agent runs on — at firm scale.

WHAT'S INSIDE

The loop, one page

Vendor-scale and firm-scale on the same architecture.

Correction Capture Log

Excel template, multi-client, encoding-ready fields.

Vertical Narrowing Worksheet

Score 3–5 workflows. The winner is your first loop.

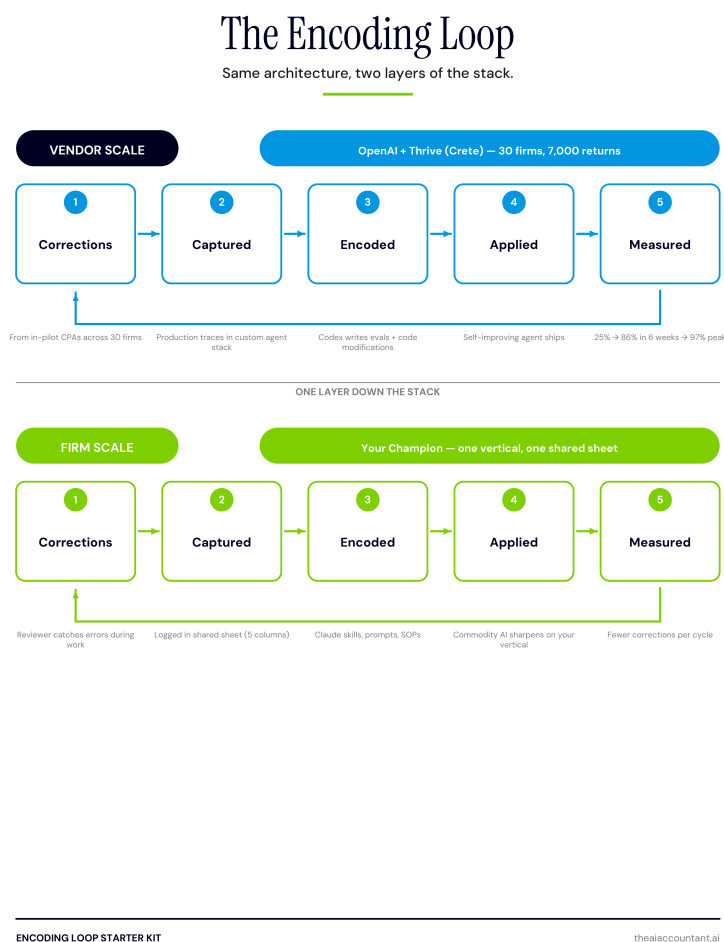
Champion role spec

One-page role definition — your forward-deployed engineer.

90-day roadmap

Three phases. Weekly checkpoints. Kill criteria.

The loop, one page



The mechanism that built Crete's 97% tax agent is structural, not technological. Corrections become captured production traces. Traces become evaluations. Evaluations become an updated agent. The cycle repeats. At Thrive's scale, OpenAI's forward-deployed engineers wrote the evaluation suites and the code modifications, and Codex shipped the self-improving updates. None of that scales down to a ten-person firm.

The same loop runs one layer down the stack. Corrections become captured rows in a shared sheet. The sheet becomes encoded firm methodology — Claude skills, prompt templates, client context files, SOPs. The encoded methodology runs against commodity AI that gets meaningfully better every month on the work you actually do. Discipline, not capital. And that's a fight your firm can pick.

Run the loop — four steps

The four steps below already exist as one-line summaries on the Start Here tab of the workbook. This page is where each step gets explained — what it looks like in practice, what good looks like, and what failure looks like.

1. Pick the vertical, not the firm.

Use the Vertical Narrowing Worksheet. Score 3–5 candidate workflows on frequency, repetition across clients, and pass/fail clarity. The single highest scorer is your first loop. The instinct to encode the whole firm is the wrong instinct. Crete's 7,000 returns weren't 7,000 different things — they were 1040s and 1041s, repeated.

2. Capture during the work, not after.

Every time the AI produces a draft and a human corrects it, add one row to the Correction Capture Log while you're still on the screen. The biggest failure mode is "I'll log it later" — you won't. Be specific in "Why the AI was wrong," "Firm-specific rule it missed," and the Pattern key. Generic categories don't produce encoded rules; specific rules and Pattern keys do.

3. Encode at the end of each week.

Filter the Capture Log by Times seen ≥ 2 (the auto-counted column on the Pattern key). Those are encoding candidates. Promote each one to the Encoding Backlog, write the rule into its target — a Claude skill, a prompt template, a client context file, or a referenced SOP — and flip the status to Live. The senior on the vertical signs off on each encoded rule before it ships. One-offs (Times seen = 1) stay in the log as evidence, not in the Backlog as work.

4. Measure monthly.

Two numbers: corrections-per-job on the chosen vertical (leading indicator — should trend down as rules accumulate) and time-on-job (lagging indicator — follows the leading indicator by 2–4 weeks). If both numbers are flat after 6 weeks, the vertical isn't producing encodable signal — re-score and consider a kill.

Pick the vertical

Volume × similarity = encodable signal. Crete had 30 firms feeding 7,000 returns of two types into one custom agent. You have one firm. You need one vertical where the volume and similarity are high enough that the loop can find a pattern within a quarter.

Three scoring questions

Frequency

How often does the firm do this workflow? Score 1 (a few times a year) to 5 (every week).

Repetition across clients

How structurally similar is this workflow across the clients you do it for? Score 1 (every client is different) to 5 (the same steps, the same rules, every time).

Pass/fail clarity

Can a senior look at the AI's draft and tell quickly whether it's right? Score 1 (judgment call every time) to 5 (objective check).

Pick the highest total. Resist running two loops at once.

The Vertical Narrowing tab in the workbook computes the total for you and includes five scored sample workflows. Replace them with yours, decide, and write the chosen vertical at the bottom of the worksheet. That's the only workflow the loop runs on for the next ninety days. The second loop is a Q2 decision, not a week-1 ambition.

Kill criteria

Re-score the vertical if: fewer than 20 correction rows by end of week 4 (no volume), fewer than three patterns showing up twice by end of week 5 (no signal), an encoded rule shows up as a correction again within two weeks of going live (rule isn't load-bearing or isn't where the AI reads it), or the Champion is spending more than 4–6 hours/week on capture (capture should be 30 seconds per row).

Capture the corrections

The Correction Capture Log is the input pipeline for everything that comes next. If the log is empty, the loop is empty. If the log is full of generic entries, the encoded rules will be generic too. The fields below are the ones that turn a passive error log into a usable encoding queue.

The load-bearing fields

Why the AI was wrong

Pick from the dropdown: Fact error, Missing context, Wrong assumption, Stale rule, Formatting, Categorization, Calculation, or Other. "Missing context" and "Stale rule" are the two that almost always produce encodable rules; the others are usually one-offs.

Firm-specific rule it missed

The one field most reviewers underwrite. Name the actual rule, not a generic category. Not "client industry treatment" — "Operating-line draws hit chequing as round-number deposits; flag any unmatched deposit \geq \$10k." The specificity here is what the encoded rule will inherit.

Pattern key (NEW in v2)

A short reusable tag for the kind of catch — "1099K-platform-fees," "operating-line-draw," "owner-sweep." The Times seen column auto-counts how many rows share the same Pattern key. ≥ 2 = ready for the Encoding Backlog. 1 = stays in the log as evidence. This is the anti-noise filter.

What to encode

Claude skill, Prompt template, SOP, Client context file, Tool config, or No fix needed. Choosing "No fix needed" is honest — sometimes the catch was a judgment call no encoded rule will replicate. Don't encode noise.

Encoded? (Y/N/In progress)

What makes the log a loop. When the rule ships in the AI tool, flip this to "Yes." Filter on "Yes" to see what you've encoded; filter on "No" to see your encoding backlog.

Encode the recurring corrections

Encoding is the work that turns the log into a moat. A correction that never makes it back into the AI's context is a correction you'll re-catch next month. The Encoding Backlog tab tracks what you've queued, what's live, and what next-cycle check confirms the encoding actually held.

Where each rule actually goes

Claude skill

A reusable, named skill file (SKILL.md) the firm's Claude project loads. Best for procedural rules that apply across many clients in the vertical — gig-1099-K reconciliation, schedule-C-from-Square-Stripe-Etsy, monthly close standard journal entries.

Prompt template

A reusable starting prompt the Champion paste-loads when running the workflow. Best for rules about how to think about a workflow — itemize-vs-standard with state-tax delta, advisory deliverable structure, partner-meeting briefing format.

Client context file

Per-client CONTEXT.md or equivalent the AI tool loads when working that client. Best for client-specific exceptions — Northwind banking rules, Summit Dental owner sweep, Chen's gig income split.

SOP

A firm-wide written procedure the AI tool references and the Champion enforces. Best for rules that need human judgment to apply — when to flag a return for partner review, when to push back on a client classification.

Measure

Two numbers, monthly. Corrections-per-job on the chosen vertical — leading indicator, should trend down. Time-on-job — lagging indicator, follows by 2–4 weeks. Both flat after 6 weeks means the vertical isn't producing encodable signal; re-score and consider a kill. Both trending down means the loop is working; open a second vertical in Q2.

Pairs with the QC Starter Kit

If you're running the QC Starter Kit at theaiaccountant.ai/qc-starter-kit, the kits work as two halves of one operating system. They aren't duplicates — they sit at different layers of the same loop and feed each other.

QC STARTER KIT

Layer 6 — Quality control

Catch what the AI got wrong on a finished piece of work. Make the checking visible. Correction Log + Exception Register, run review-by-review.

Output: a structured record of catches and the standing rules that should be enforced going forward.

ENCODING LOOP STARTER KIT

Layer 4 — Encoded methodology

Turn those catches into firm methodology the AI will actually use next month. Correction Capture + Vertical Narrowing + Encoding Backlog.

Output: encoded Claude skills, prompts, SOPs, and client context files that make the same correction stop recurring.

How they connect

Each row in the QC Kit's Correction Log is a candidate row for this kit's Correction Capture Log. The QC Kit answers “did the AI get this one right?” — review by review. The Encoding Loop Kit answers “why does the AI keep getting this kind of thing wrong?” — month by month. Together: the firm catches errors AND the errors stop recurring. That's the operating system. Two surfaces, one loop.

The first 90 days

The point of the first quarter isn't to encode the whole firm. It's to prove the loop runs on one vertical, encode the highest-frequency missing rules, and finish with a Champion who defends the role with numbers.

DAYS 1–30 Pick the vertical, set up capture

Pick the vertical with the Narrowing Worksheet. Open the Capture Log in a shared location every reviewer can reach in two clicks. Tag every row with a Pattern key. Don't encode yet — gather data first. EXIT: ≥ 40 log rows, ≥ 3 Pattern keys with Times seen ≥ 2 , 1 named vertical.

DAYS 31–60 Run the loop, accumulate corrections

Filter by Times seen ≥ 2 . Promote the top three Pattern keys to the Backlog. Encode them. Senior reviews each before it ships. Watch the same workflow run again in week 8 — encoded rules should stick. EXIT: 3–6 rules live, 80–150 cumulative rows, ≥ 1 verifiable recurrence prevented.

DAYS 61–90 Encode the rest, measure, defend the role

Encode the next batch. Pull two numbers: corrections-per-job and time-on-job, week 1 vs. week 11. Write a one-page report. Present to the owner. EXIT: 8–15 rules live, corrections-per-job down $\geq 25\%$, 1 named second vertical for Q2.

What you finish the quarter with

Not a 97% tax agent. Something different and harder to copy. A documented Correction Capture Log against one vertical. Eight to fifteen encoded rules competitors can't replicate by buying the same AI tool. A measurable drop in corrections-per-job on the chosen workflow. A Champion who defends the role with numbers. And one named second vertical for Q2.