6 guardrails

for letting an AI agent
touch a live ad account

Claude can already draft campaigns and push them live. These are the rules I want in place before it can spend.

01 Plan before launch

02 Read the ad back

03 Arm auto-pause rules

04 Audit for waste weekly

05 Keep a change log

06 Test incrementality

synterai.com

Plan before launch

WHY IT MATTERS

I do not let an agent publish anything until I have approved a written plan.

HOW TO DO IT

1. Have Claude write the full campaign plan first, including the objective, audience, budget, creative, and destination URL.
2. Read every line yourself and mark what needs to change.
3. Approve the plan in writing. Only then can anything go live.

GIVE CLAUDE

Write a complete campaign plan for [account]. Include the objective, audience, budget, creative, and destination URL. Do not launch anything until I approve this plan in writing.

Read the ad back

WHY IT MATTERS

The ad that shipped is not always the ad I approved, so I check the live version on the platform.

HOW TO DO IT

1. After launch, pull each live ad from the platform itself, not from Claude's notes.
2. Check the creative, the button people click (the CTA), and the URL against the approved plan.
3. If anything does not match, fix it before more budget goes out.

GIVE CLAUDE

Pull every live ad in [campaign] from the platform. For each one, show the creative, the button people click, and the destination URL, compare them to the approved plan, and flag anything that does not match.

Arm auto-pause rules

WHY IT MATTERS

I write the stop rule before the campaign starts so an ad set cannot keep spending with no results.

HOW TO DO IT

1. Set a minimum age and a spend threshold for each ad set. An example setting: 3 days and 2x your target CPA.
2. If an ad set hits those limits with zero conversions (no results), pause it and send a notification.
3. Review paused ad sets before turning them back on.

This pauses the ad set. It is not a hard cap on total spend.

GIVE CLAUDE

Set an auto-pause rule on [campaign]. Example setting: if an ad set spends more than 2x the target CPA of $[X] over 3 days with zero conversions (no results), pause it and tell me. Do not change the rule without my approval.

Audit for waste weekly

WHY IT MATTERS

Weak placements and overlapping audiences can keep spending unless I look for them on a schedule.

HOW TO DO IT

1. Once a week, look for weak placements, overlapping audiences, and budget that should move.
2. Have Claude propose each change with a reason.
3. You decide what to apply. Claude does not edit the account on its own.

GIVE CLAUDE

Audit [account] for waste over the last 7 days. Find weak placements, overlapping audiences, and budget that should move. Propose each change with a reason. Do not apply anything until I approve.

Keep a change log

WHY IT MATTERS

If I cannot explain a change in the account,
I do not leave it in place.

HOW TO DO IT

1. Log every edit with who made it, what changed, and when.
2. Read the log before you raise budgets or debug a problem.
3. If a change has no explanation, roll it back.

Google Ads has a native change history.
On other platforms, keep your own log.

GIVE CLAUDE

Show the change history for [campaign] over the last [N] days.
For each edit, list who made it, what changed, and when.
Flag anything I cannot explain, and do not make further edits
until I review.

Test incrementality

Check whether the ads actually caused the sales.

WHY IT MATTERS

The ROAS a platform reports (return on ad spend) will include sales I would have gotten anyway.

HOW TO DO IT

1. Before you scale a channel, run a holdout: a geo or audience group that does not see the ads.
2. Compare conversion lift in the group that saw the ads against the holdout.
3. Raise budget only if the holdout shows real lift, not just because platform ROAS looks strong.

GIVE CLAUDE

Design a holdout test for [channel] before we scale spend.
Propose a geo or audience group that will not see the ads, a test window, and how we will measure real lift versus the ROAS the platform reports.
Do not change budgets until I approve the design.

Each card has a paste-ready prompt you can give Claude.

I will drop the six skill files in the first comment.
If you already use guardrails on agents that can spend,
which ones do you require before anything goes live?