

AGENT GUARDRAILS

My ads tool cannot turn an ad on

That is on purpose. Three guards make sure of it.

Agents draft. Humans press the button.

Any ad my agent creates starts paused.

Guard 1: paused only

A request that is not paused is refused before it is sent.

Guard 2: a budget ceiling

A daily budget over 50 dollars is refused.

Guard 3: no deletes

Deletes and archives never come from a chat.



The guard tests itself

11

SELF-TEST CASES

11 of 11

PASSED

11 October 2026

READ
ON

Build your own guard today

- 01 List everything your agent can write to
- 02 Make the safe state the default: paused, draft
- 03 Add a script that refuses anything else
- 04 Test it on a case that should fail



The cost of waiting for me is 30 seconds.

— DONAL, ON WHY THE GUARD EXISTS



Comment FLEET

I'll send the agent guardrails list