

M A V E R I C K A I

The Canary Trick

A ONE-LINE EARLY WARNING FOR WHEN CLAUDE STARTS TO DRIFT

The longer a chat goes, the more an AI starts to forget, drop your instructions, and make things up, and it never warns you. The fix is one tiny instruction that acts like an early warning sign. The moment it disappears, you know it's time for a fresh chat. Takes 30 seconds to set up.

Why long chats quietly start to degrade

What a 'canary' is and why it works

The one line that sets yours up in 30 seconds

Exactly how to read it (and what to do when it drops)

More canary types beyond your name

A one-page cheat sheet

2026 Edition

Why Long Chats Go Bad

Every AI chat has a **context window**, basically, how much of the conversation it can hold in its head at once. While you're inside that window, it's sharp. But a long conversation slowly fills it up, and once you push past it, the model starts quietly dropping things: it forgets earlier details, ignores instructions you gave it, and fills the gaps by making things up. That's where the confident, wrong answers come from.

The real problem isn't that this happens, it's that the AI never tells you it's happening. It sounds just as confident on turn 80 as it did on turn 3. So you need your own early warning sign. That's the canary.

WHERE THE NAME COMES FROM

Miners used to carry a canary down into the coal mine. The bird was more sensitive to bad air than they were, so if it stopped singing, they knew to get out before they were in danger. Your AI canary works the same way: a small signal that fails first, so you know to act before the answers go bad.

Set Up Your Canary (30 Seconds)

Your canary is one simple standing instruction: have the AI start every reply with your name. It's a tiny, easy task, so it's the first thing to slip when the context starts degrading. When the name vanishes, that's your bird going quiet.

STEP 1: Give it the instruction

In Claude, paste this and send it:

THE CANARY

```
From now on, start every single response with my name,  
[YOUR NAME], and save this instruction to your memory  
so it applies going forward.
```

STEP 2: Make it stick

Saving it to memory makes it apply across chats. Even better, put the same line in any standing instructions you use, your personal preferences / custom instructions, a project's instructions, or an instructions file if you work with those. That way your canary is always on, in every conversation.

TIP

Use your actual first name, something you'll instantly notice is missing. "Hey Sarah, here's..." at the top of every reply becomes so normal that its absence jumps out at you immediately.

How To Read Your Canary

Now you just watch for one thing while you work.

What you see	What it means
It's still using your name	The bird's singing. Context is healthy, keep going.
It stops using your name	The bird went quiet. The context is degrading, time for a fresh chat.

The moment Claude drops your name, treat it as your cue: this session has gotten long enough that quality is slipping, so start a new chat before you rely on anything else it says.

IT'S AN EARLY WARNING, NOT A LIE DETECTOR

Be fair to the tool here: the canary dropping doesn't mean everything the AI just said is false, it means quality MAY be slipping and it's smart to refresh. And once in a while it might drop the name for an unrelated reason, so it's the name consistently disappearing that's the real signal. Either way, when in doubt, a fresh chat costs you nothing.

When It Triggers: The Clean Handoff

Starting a new chat doesn't mean starting over. Before you open a fresh one, have the current chat pack everything important into a summary you can carry across.

HAND OFF TO A FRESH CHAT

```
Before I start a new chat: summarize everything that  
matters from this conversation, our goal, the key  
decisions we made, where we left off, and the exact  
next steps, so I can paste it into a fresh chat and  
pick up right where we are.
```

Paste that summary as the first message in your new chat (and re-add your canary line), and you're back to a sharp, healthy context with none of the drift, but all of the progress.

THIS IS THE WHOLE LOOP

Work in a chat, watch the canary, and the moment it drops: summarize, open a fresh chat, paste the summary, re-set the canary, keep going. Do that and you basically never get burned by a degraded session again.

More Canaries (Once You're Hooked)

Your name is the easiest canary, but any small, consistent instruction works. Pick whatever you'll notice, or layer a couple for extra warning.

Canary	The instruction	It fails when...
The name	Start every reply with my name	Your name disappears
The tag	Begin every reply with the tag [ON TRACK]	The tag stops showing up
The counter	Number each reply (Turn 1, Turn 2...)	The count skips or stops
The goal line	End each reply with 'Goal: [my goal]'	It forgets or changes the goal

TIP

For important, long work, use two canaries at once, say, your name at the top and your goal at the bottom. If either one slips, you get an even earlier warning that it's drifting.

The One-Page Cheat Sheet

The whole trick, in order. Screenshot this.

1. SET THE CANARY

From now on, start every response with my name,
[YOUR NAME], and save this to your memory.

2. WHEN THE NAME DROPS, HAND OFF

Summarize everything important, goal, key decisions,
where we left off, next steps, so I can paste it into
a fresh chat.

The rules that make it work

Put the canary in your saved memory or instructions so it's always on. Use your real name so you notice instantly. When it drops, it's a signal to refresh, not proof of a lie. Summarize and carry over so you never lose progress. And re-set the canary in the new chat.

YOUR ACTION PLAN

1. Add the canary line to Claude's memory / your instructions right now. 2. Work like normal and keep half an eye on your name at the top. 3. The moment it disappears, run the summary prompt. 4. Open a fresh chat, paste the summary, re-add the canary. 5. Never get burned by a quietly degraded chat again.

Want More AI Tips Like This?

I post daily AI content, tool breakdowns, and simple tricks that actually make your AI more reliable.
Follow me for the latest.

@MavGPT

TikTok | Instagram | Facebook

When the canary goes quiet, start a fresh chat. Simple as that.