CARE

WHITEPAPER

# When Two Instructions Conflict, Who Wins?

**Explicit escalation keeps uncertain authority from becoming arbitrary execution.**

Control begins with a clear distinction between preparation, permission and accountable human decision.

**AGENTIC CARE**

https://ai-managementsystem.com

## Contents


01 The agent should not invent a hierarchy

02 State the stopping condition

03 Escalation is a process step

04 Make authority testable

05 What management can verify

06 Start with one agent

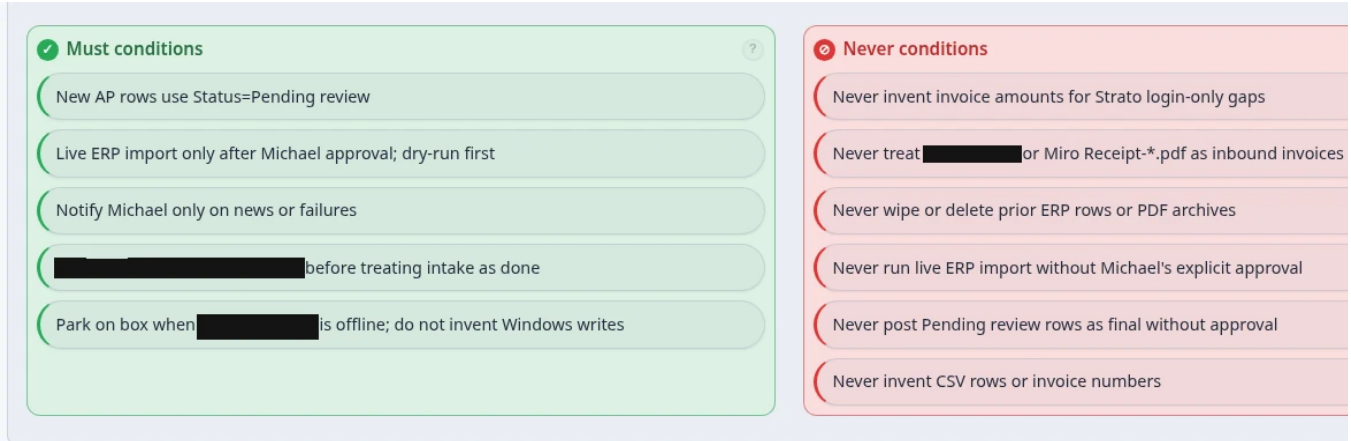

## The agent should not invent a hierarchy

An owner may place an invoice on hold while a coordinating agent sends a handoff requesting import. Both messa operationally credible. The AP example records a real conflict of this kind. Automatically following whichever instru not a safe management policy.

> **KEY FINDING**
>
> Explicit escalation keeps uncertain authority from becoming arbitrary execution.

## State the stopping condition

The CARE action list proposes that when an owner's direct message conflicts with a relayed instruction, the importi asks which applies. The Never conditions prohibit treating approval quoted by another bot as owner approval. This conversational problem into a clear operating boundary.



A useful procedure identifies the trigger, recipient and required evidence. It should also say what remains on hold a non-conflicting work can continue. Without these details, escalation becomes an abstract promise that collapses ur

## Make authority testable

The organization can evaluate the agent with contradictory messages and confirm whether it refrains from import. of behavior under the tested conditions, not a universal guarantee. CARE's accepted definition identifies what the b and who may change it.

> **MANAGEMENT IMPLICATION**
>
> Management takeaway: retain the decision and verify the operating evidence before treating a delivered control.

## What management can verify

The following checkpoints translate the paper's argument into a review of the actual operating arrangement. They a not claims that a proposed control has already been implemented.

| Checkpoint | Management question | Evidence to |
|---|---|---|
| **Conflict** | Direct owner hold contradicts coordinator request | Conflict dete |
| **Pause** | No live write while unresolved | Safe stop co |
| **Escalation** | Ask authorized owner for explicit resolution | Decision attr |

Where evidence is missing, record the gap and the accountable owner rather than treating the document itself as p

## Start with one agent

Use a real agent and review what it produces.

1. the agent to https://ai-managementsystem.com/under-care.
2. it is finished, log into CARE and review its documentation.

**AGENTIC CARE**

Instant AI Management System for your Agents, as a Managed Service.

**https://ai-managementsystem.com**