

A person at the helm

How a person authorises irreversible actions

Who decides before AI sends something?

A person decides on content, recipient and action. AI can prepare a proposal; fixed checks can report findings. Neither replaces that decision. If the proposal changes, review the decision again.

Summary

- **The question:** how do you stay in control when AI prepares an action with consequences beyond your workspace?
- **What we did:** we turned public working principles into a fictional exercise about a quotation.
- **What we saw:** preparation, checking and deciding each have a role. We measured no improvement in safety.
- **What it means:** approval belongs to a specific proposal that you can inspect.
- **The main limitation:** a human decision can be wrong. This exercise does not demonstrate a safe implementation.

Approval belongs to a proposal

Before approval, put the content, recipient and version beside the decision. You can then assess what will actually happen.

A fingerprint does not make an amount or recipient correct. You still need to inspect the proposal.

Try a fictional quotation

This is a fictional exercise. The rule: a person sends after checking the content and recipient. Findings must be resolved first. Choose what follows that rule.

What do you do with this quotation?

Step	Situation	What happened	Method
1	An AI has prepared a quotation. You have not checked its content or recipient yet.	In this exercise the quotation stays inside. You inspect the proposal first, then decide about sending it.	Inspect first
2	You approved it. The recipient address then changed. You have not reviewed this version.	Here, the earlier approval does not cover the changed proposal. You review again what will go to whom.	Decide again
3	The check now reports an internal reference. The agreed rule requires resolving that finding first.	The quotation stays inside. After correction, the check and the decision are repeated for this version.	Resolve first

Inspect the proposal, resolve findings and decide on the version that will actually leave.

A check and a person do different work

A fixed check recognises agreed patterns. A person also assesses purpose and context. Passing a check does not prove an action is desirable. An earlier approval does not replace assessing new findings.

Method

We made three exercise steps from public rules on human approval and the gate for outgoing work. Each correct choice follows from the rule stated beforehand. We ran no model test and sent nothing.

What this evidence says

The working policy describes a method. The page *Our commitments* also sets research goals for data, model training, services and shared lessons. Those goals remain research goals. A written rule does not prove that every implementation follows it.

Limitations

- **Not representative.** ITARR's workshop does not represent small and medium-sized businesses.
- **Fictional exercise.** No real quotations, clients or outcomes were used.
- **No guarantee.** People and checks can miss mistakes.
- **No implementation manual.** Who may decide, and which checks are needed, must be defined for each application.

What this means

Take one action your AI prepares. Before approval, can you see exactly what will happen, for whom and why? Start with that specific question.

Closing

With thanks to readers who examine the exercise critically.

How to cite: ITARR Lab (2026). *A person at the helm*. Version 0.1, reference date 25-10-2026. itarr.eu/lab/de-mens-aan-het-roer. CC BY 4.0.

Sources: [Regie over je AI](#), especially laws 1, 2 and 4; [Onze afspraken](#). Both consulted on reference date 25-10-2026.

Change log: 0.1 (25-10-2026): first draft with a fictional exercise.

Status: draft for review; not yet approved by a person.

Licence: CC BY 4.0. Credit ITARR Lab, link to the licence and indicate changes.