

Stay in control of your AI

A working policy for AI that does real work

Contents

Summary	1
Foreword: sharing multiplies knowledge	1
The laws of the workshop	2
How to start tomorrow	5
Limitations	5
Open questions	5
What comes next	5
Closing notes	5

Summary

- **The question:** how do you let AI do real work in an ordinary business without losing control?
- **What this is:** the rules under which ITARR works, with three people and many AI colleagues.
- **The core:** a person approves irreversible actions. Everything leaving the workshop passes through one gate. Sensitive data follows its own route.
- **Where it comes from:** our own mistakes, including a night with 20,478 failed tasks that nobody noticed.
- **The main limitation:** we have not finished implementing every law ourselves, our workshop is not representative, and this is not legal advice.

Foreword: sharing multiplies knowledge

At ITARR, AI does real work. AI colleagues write, investigate, check and build. An AI colleague, also called an agent, is an AI that takes steps by itself. We have three people and many AI colleagues. That works only with clear agreements. What may an AI colleague do on its own? Where does a person approve? What must never simply be sent out?

Those agreements did not come from a book. We learned them, often because something went wrong. One night in September, two of our systems got stuck in a loop. One kept requesting capacity. The other kept saying yes. By morning, there were 20,478 failed tasks. Other work had been pushed aside 1,533 times. Nobody had seen it, because nobody was awake. Nights like that produced the rules in this article.

We share them for three reasons.

- Knowledge does not run out when you share it. We keep what we know, and you gain it too.
- More and more businesses are starting to work with AI. Better to begin with sound agreements than expensive lessons.
- People who use it can improve it. We would welcome what you learn for the next version.

What you may do. Use, adapt and share everything, including within your own business and for your clients. This article is licensed under CC BY 4.0. Attribute the source, link to the licence and indicate any changes.

What we never share. Client data, client documents or anything that identifies a client. Nor do we share how our own systems and keys are set up. This article is about how we work, not whom we work for.

An honest note. We also use AI from US providers, including for writing, programming and processing documents. Alongside that, we run open models on our own machines: AI models that people can download and run themselves. Staying in control does not mean doing everything yourself. It means deciding, task by task, where your data goes. And knowing why.

The laws of the workshop

These are the rules under which our AI colleagues work. Each law explains why it exists and how to apply it yourself. Under Our practice, we state where we stand, including what is not finished.

Law 1. A person approves irreversible actions

An AI colleague may do ordinary work by itself. It may also do consequential work, but with a log. Actions that cannot be undone wait for a person's approval.

Why. An AI colleague works quickly and does not get tired. That is precisely the risk: an error travels just as fast as good work. Sending, paying, deleting and publishing cannot simply be taken back.

How to apply it.

- Divide your work into three levels: ordinary, consequential and irreversible. Do this in advance for each type of task.
- Before every consequential step, ask four questions. What am I changing? Why? What if it goes wrong? Is this allowed?
- Make approval a deliberate act, rather than a tick at the bottom of a long list.

Our practice. A fixed rule. Irreversible steps wait for explicit approval from our founder. Since October 2026, that can also be given with a fingerprint.

Law 2. A repair is a new change

Did something go wrong, and are you going to fix it? The repair is itself another change, with the same four questions and the same level.

Why. Rushing causes the second error. Someone making a quick repair often skips the very check that would have caught the first mistake.

How to apply it.

- Treat every repair as a new task, with its own entry in the log.
- If the repair is irreversible, a person approves it again.

Our practice. A fixed rule in our working framework.

Law 3. One retry, then a person

Two systems triggering each other without a brake can fill everything up in one night. An automatic attempt therefore gets one retry. If it fails again, a person takes a look.

Why. That night in September, one system repeatedly requested capacity and another kept saying yes. Our monitor saw the machine getting warm. But it had no rule for the number of

tasks. By morning, the counter had reached 20,478. We have no baseline: at the time, we were not yet counting failed tasks on ordinary nights.

How to apply it.

- Give every automatic repetition a limit: one retry.
- Have a monitor count events as well as measure temperature or detect faults.
- Provide an emergency brake a person can use in one action.

Our practice. The rule has existed since September 2026. The system involved remains switched off until it is rebuilt with this brake.

Law 4. The outward gate is narrow

Everything leaving the business passes through one gate: an email, a quotation, a report or a post on the internet.

Why. An AI colleague picks up words from everything it has read. An internal working name, a client name or a file location can then slip into a document. That happened to us too.

How to apply it.

- An AI colleague prepares drafts. A person sends them. Always.
- Before anything goes out, scan it for internal names, client names, locations and keys.
- If there is doubt, it stays inside.

Our practice. A person sends messages. Since October 2026, each draft we prepare goes through a scan first. If it finds an internal trace, the draft is not created. An exception requires a recorded reason.

Law 5. Sensitive data follows its own route

Not every AI service may see everything. You decide where each type of data may go.

Why. A question about a public text is different from a client file. Yet many businesses send everything through the same service. That is how you lose control of your data without noticing.

How to apply it.

- Make a short list of the permitted route for each type of data. For example: client files only on your own machines or at a European service; public texts may go anywhere.
- Remove names and numbers before data leaves your organisation, where possible.
- Provide a way out for each service. How do you switch if you need to?

Our practice. Not finished. Some of our client work runs on our own machines. For other work, including work with client documents, we use a US service. We are still preparing a complete list for each type of data.

Law 6. Record what was decided, not who said what

What is not written down will not be there tomorrow. Record the goal, what was agreed and what remains open.

Why. An AI colleague remembers nothing from yesterday unless it is recorded somewhere. A person returning from holiday can be much the same. People also speak more freely when they know their words will not be reproduced literally.

How to apply it.

- Give every decision a reason, in a consistent place.

- Write reports as goals, agreements and unresolved questions. No quotations unless someone explicitly asks for them.
- After a mistake, put the lesson alongside the rules so everyone reads it too.

Our practice. We record decisions and lessons in a consistent place. Reporting without quotations has been a rule since 16 October 2026.

Law 7. A second look, from someone else

Each piece of work gets checked by a fresh reviewer. Preferably a different AI model from the one that created it.

Why. The same model has the same blind spots. It reads past its own mistake. A reviewer forced to find something may also start inventing faults.

How to apply it.

- Give the reviewer a short brief: the requirements, the acceptance criteria and where the work is.
- Ask only for findings supported by evidence.
- Say explicitly that finding nothing is a valid answer.

Our practice. Since September 2026, every delivery of a build gets a second look from another model.

Law 8. Trust grows step by step

A new AI colleague does everything through a person. A proven colleague works independently, with a log. Only routine work runs automatically.

Why. Trust is earned through work that went well. That applies to people, and it applies here too.

How to apply it.

- Start each new AI task as a proposal: the AI proposes, a person decides.
- Let a task run independently only after a series of successful rounds.
- Did something go wrong? Move the task back a level.

Our practice. This is our approach, but we are still defining the fixed levels.

Law 9. Client work comes first

Computing capacity, time and money are finite. When capacity runs out, stop your own experiment first. Never the client's work.

Why. Every step costs something: electricity, money or waiting time. Value comes only from work that has been delivered and accepted.

How to apply it.

- Define in advance what takes priority when capacity runs out.
- Give every task a budget and check its actual cost.
- Choose the smallest model that does the task well.

Our practice. Our own machines are configured so that, when memory fills up, our own experimental work stops first.

How to start tomorrow

1. Write down your three levels: what is ordinary, consequential and irreversible in your organisation?
2. Assign one person to approve irreversible actions.
3. Make the list from law 5: which data may go where?

The other laws follow from there.

Limitations

- **Not representative.** Our workshop is small: three people and many AI colleagues. Your experience may differ.
- **Not fully implemented here.** Laws 5 and 8 are not yet fully in place in our workshop. Their sections say so.
- **Not legal advice.** Where we mention regulations, we give a source and reference date. If in doubt, ask a lawyer.
- **Not finished.** This is version 0.1. A fillable policy and a guide connecting it to regulations will follow.
- **Not a ready-made package.** Implementation differs from one business to another.

Open questions

- Which law is hardest in your organisation, and why?
- How do you make law 5 manageable for a business without its own IT department?
- When does an AI colleague earn another level of trust?

We would like to hear what you observe.

What comes next

- **A fillable policy template:** levels, gate, data routes, checks and events.
- **The connection with regulations:** what the AI Act requires, with sources and reference dates.

Also read in ITARR Lab: the film of the September night and the lesson on how an AI colleague catches its own mistakes.

Closing notes

With thanks to the people and AI colleagues at ITARR who work with these rules and improve them.

How to cite this: ITARR Lab (2026). *Stay in control of your AI*. Version 0.1, as of 17 October 2026. itarr.eu/lab/regie-over-je-ai. CC BY 4.0.

Sources: ITARR's own working practice, as of 17 October 2026; a night in September 2026, shown in the film at itarr.nl/etalage/nacht/ (as of 17 October 2026).

Change log: 0.1 (17 October 2026): first draft.

Licence: CC BY 4.0. Use, adapt and share it. Attribute ITARR Lab, link to the licence and indicate any changes.