

REINHOLD SCHILLING

GxP / AI Governance

*From interesting tools
to controlled processes*

17th International SLS Days · 24 September 2026

Wörwag Pharma GmbH & Co. KG

THE SITUATION

The adoption gap nobody reports

AI arrived in our companies through the side door, not through a project plan.

INTERESTING TOOLS

Everywhere

Drafting, summarising, translating, screening, coding.

Adopted individually, quickly, and mostly without anyone deciding.

CONTROLLED PROCESSES

Almost nowhere

Defined purpose, named owner, documented limits, evidence on file.

Rarely in place, including in regulated workflows.

The distance between those two boxes is where the audit finding lives.

STANDARD ANSWER 1

“We validated the tool”

Validation certifies a system. It does not certify what you do with it.

01

The output is not reproducible

Same prompt, same document, different answer. Classic qualification logic assumes it will not do that.

02

The system changes without you

Models, prompts and vendor settings move between releases you never approved.

03

It says nothing about the use case

Safe for polishing an email does not mean safe for drafting a regulated record.

Useful and necessary, but it answers a question nobody is really asking.

STANDARD ANSWER 2

“A human checks it anyway”

Human review only works if the error is visible in the output.

REVIEWERS RELIABLY CATCH

Visible problems

- Statements that are obviously wrong
- Language that sounds off
- Missing sections you expected to see

REVIEWERS ROUTINELY MISS

Invisible changes

- Something quietly left out
- “must” turned into “should”
- A condition or exception dropped
- A responsibility shifted

If a qualified reviewer cannot spot the error without returning to the source, human-in-the-loop is a label, not a control.

THE REFRAME

Stop assessing tools. Assess uses.

Same model, same vendor, same week, two completely different risk situations.

USE A

Improving the wording of a draft

- Human wrote the content
- Source is in the room
- Error is visible on the page
- Nothing regulated is created

Light controls. Get on with it.

USE B

Turning sources into a regulated record

- Machine selected what matters
- Sources are not re-read
- Omissions are invisible
- Output becomes the record

Full assessment before go-live.

The unit of governance is the tool in a specific context of use.

WHAT IT TAKES

Five things that carry the weight

Not a control catalogue. The elements without which the rest is decoration.

- 1 A named owner**
One person accountable for this use.
- 2 A defined context of use**
What goes in, what comes out, what the output may become.
- 3 Non-negotiable gates**
Conditions no other controls can compensate for.
- 4 Evidence sized to risk**
A screenshot for low risk, a documented evaluation for high risk.
- 5 Benefit after correction**
The net result once review and rework are counted.

Number five is almost never measured.

THE MEASURE THAT IS MISSING

Human Correction Burden

Everybody measures how good the model is. Almost nobody measures what is left afterwards.

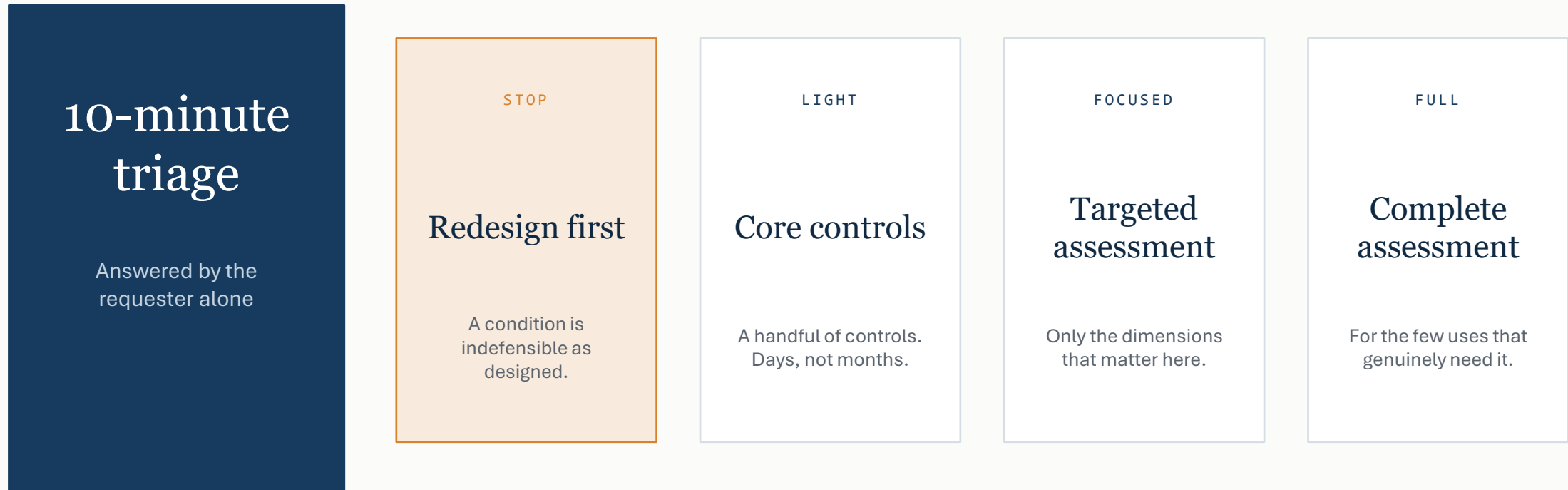
$$\begin{array}{ccccc} \text{Time the tool} & & \text{Time spent reviewing,} & & \\ \text{appears to save} & - & \text{correcting, escalating} & = & \text{Real benefit} \end{array}$$

If the reviewer has to check everything anyway, the tool did not save the work. It moved the work and made it less visible.

MAKING IT WORKABLE

Governance must not become the bottleneck

Most uses are low risk. Spend assessment effort where it changes something.



Most requests land in the first two outcomes. That is exactly the point.

Interesting tools impress. Controlled processes scale.

- 01 Assess uses, not tools
- 02 Ask whether the error would even be visible
- 03 Measure what is left after correction

I am building this into a working framework and looking for reviewers and pilot partners.

Reinhold Schilling
Head of Global Pharmacovigilance
& EU-QPPV

Backup slides

- 01 What actually applies, and when
- 02 Where to start, and what to ask a vendor
- 03 How to measure it, and who owns it

*Six slides. Shown
only if the question
comes up.*

Reinhold Schilling
17th International SLS Days

IF SOMEONE ASKS ABOUT THE AI ACT

The delay is real, but narrow

Digital Omnibus in force since 27 July 2026. Verify the status before quoting dates.

What moved

and what did not. Only the heaviest regime was postponed.

2 AUG 2026

Art. 50 duties

Art. 50 duties live. Art. 4 AI literacy already since Feb 2025.

2 DEC 2026

Legacy systems

Art. 50(2) for systems already on the market. New prohibitions.

2 DEC 2027

High-risk, Annex III

Moved from Aug 2026. Stand-alone high-risk systems.

2 AUG 2028

High-risk, Annex I

AI embedded in regulated products, e.g. medical devices.

None of this replaces GxP. Annex 11, GVP and your PSMF apply today, unchanged.

IF SOMEONE ASKS WHERE TO START

Monday morning, before any policy

Three moves. Together they take an afternoon, and almost nobody has done them.

01

List what is actually used

Ask people, do not send a survey. You are looking for the tools nobody registered, which is most of them.

02

Ask what it becomes

Draft, decision, or record. Only the third category needs real work, and it is usually a short list.

03

Name one owner per use

A person, not a function. This single step resolves more than any policy document you could write first.

A policy written before the inventory describes a company you do not have.

IF SOMEONE ASKS ABOUT VENDORS

Three questions worth more than a certificate

“Validated and compliant” describes their system in their environment for their intended use.

01

What changes, and how will we know?

Model versions, system prompts, settings. Ask for the change notification commitment in writing.

02

What happens to our data?

Storage, retention, sub-processors, and whether anything we send is used to improve their service.

03

What evidence exists for our input?

Not the demo set. Our document types, our languages, our edge cases. Usually the question that stops the conversation.

If question three has no answer, generating that evidence is your cost. Put it in the business case.

IF SOMEONE ASKS FOR AN EXAMPLE

What an invisible change looks like

A summary of a procedure. Fluent, plausible, and wrong in four places.

WHAT THE SOURCE SAID

The procedure

- The QP must approve before release
- Deviations closed within 30 days
- Except for batches under quarantine hold
- Escalation to the QPPV is mandatory

WHAT THE SUMMARY SAID

The AI output

- The QP should approve before release
- Deviations are closed promptly
- (exception not mentioned)
- Escalation to the QPPV is recommended

Four changes. None of them looks like an error on the page.

IF SOMEONE ASKS HOW TO MEASURE IT

Ten cases, two people, one afternoon

You do not need a study. You need enough cases to see the shape.

1 Use real cases

Ten of them. Not demo material, and not only the easy ones.

2 Split them

Five with the tool, five without. Same reviewer, same standard.

3 Record two numbers

Minutes spent reviewing, and what actually had to be changed.

4 Read the spread

Not the average. A few full reconstructions eat the savings from all the rest.

5 Repeat after changes

A new model version or prompt makes the number stale. It is not measured once.

The spread is the finding, not the mean.

IF SOMEONE ASKS WHO OWNS THIS

Five roles, and the gap between them

Every one of these functions covers part of the problem. None covers the whole.

- | | | |
|---|----------------------|---|
| 1 | QP / Stufenplanb. | Product and safety decisions. No mandate over how tools are used. |
| 2 | QPPV | Oversight of the PV system. No visibility of AI in other functions. |
| 3 | Data protection | Lawfulness of the data use. Says nothing about output quality. |
| 4 | Information security | Confidentiality and access. Correctness is not its question. |
| 5 | CSV / QA | Validates the system. Does not own the context of use. |

The gap between them is the governance gap.