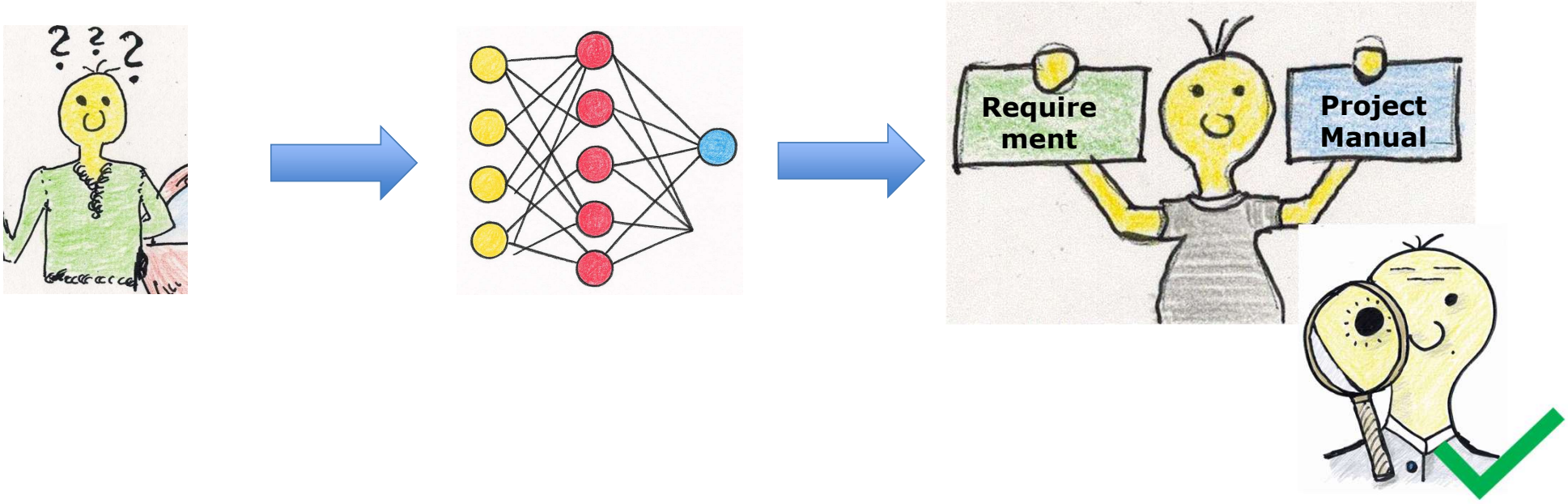
Webinar – Make it or Fake it

Human vs. AI: Efficiency in Creation, Review Effectiveness, and Fake-Evidence Detection

Learn to Fake It!



- Is it possible to fake evidence that look relevant for your project? → Yes, of course
- Can you convince Assessors that your evidence is real project documentation → Maybe
- Does this add value to your project? → Definitely not

What are our Goals today?

- Look back at the **VDA SYS workshop** and its exercises
- Compare **AI-supported and human approaches** for creating and reviewing work products
- Discuss how to recognize **weak, polished, or potentially fake evidence**
- Share the **main observations and lessons learned**
- Reflect on what this means for **engineering practice, reviews, and assessments**

Exercise: Which slide looks AI created?

Snippet 1: Excerpt from a System Requirement

ID	System Requirement	Linked Stakeholder Requirement
SYS-TG-042	During powered opening, the tailgate control unit shall regulate the actuator movement speed within ±5% of the configured nominal opening speed after the initial acceleration phase and before entering the end-position damping phase.	STRS.5

Field	Content
Rationale	The requirement refines the stakeholder expectation for constant opening speed and separates the control-relevant movement phase from transient start and stop behavior.
Verification Method	System test on actuator bench
Status	Reviewed
Review Comment	Wording aligned with system terminology. No open technical issues.

Field	Content
Rationale	The requirement refines the stakeholder expectation for constant opening speed and separates the control-relevant movement phase from transient start and stop behavior.

Snippet 3: Excerpt from a Test Summary

TSR-SYS-031 | TC-SYS-117 | SYS-TG-042
Test Object: Tailgate actuator system bench
SW / Calibration: SW_TG_2.3.1 / CAL_TG_R2_2025-10-06
Result: Passed with observation

Summary

- Five powered opening cycles were executed on the actuator system bench.
- The evaluated movement phase was defined from **1.0 s after movement start** until **1.5 s before fully open position**.
- Mean opening speed across all runs was between **118.9 mm/s and 120.7 mm/s**, within the accepted range of **114 mm/s to 126 mm/s**.
- Run 4 showed a short speed drop to **113.6 mm/s for 40 ms**. The observation was recorded as **OBS-SYS-089** and linked to the raw measurement file and bench configuration.

- Five powered opening cycles were executed on the actuator system bench.
- The evaluated movement phase was defined from **1.0 s after movement start** until **1.5 s before fully open position**.
- Mean opening speed across all runs was between **118.9 mm/s and 120.7 mm/s**, within the accepted range of **114 mm/s to 126 mm/s**.
- Run 4 showed a short speed drop to **113.6 mm/s for 40 ms**. The observation was recorded as **OBS-SYS-089** and linked to the raw measurement file and bench configuration.

Workshop at VDA SYS Conference

Workshop – Make it or Fake it

Human vs. AI: Review Effectiveness, and Fake-Evidence Detection

Workshop Objectives

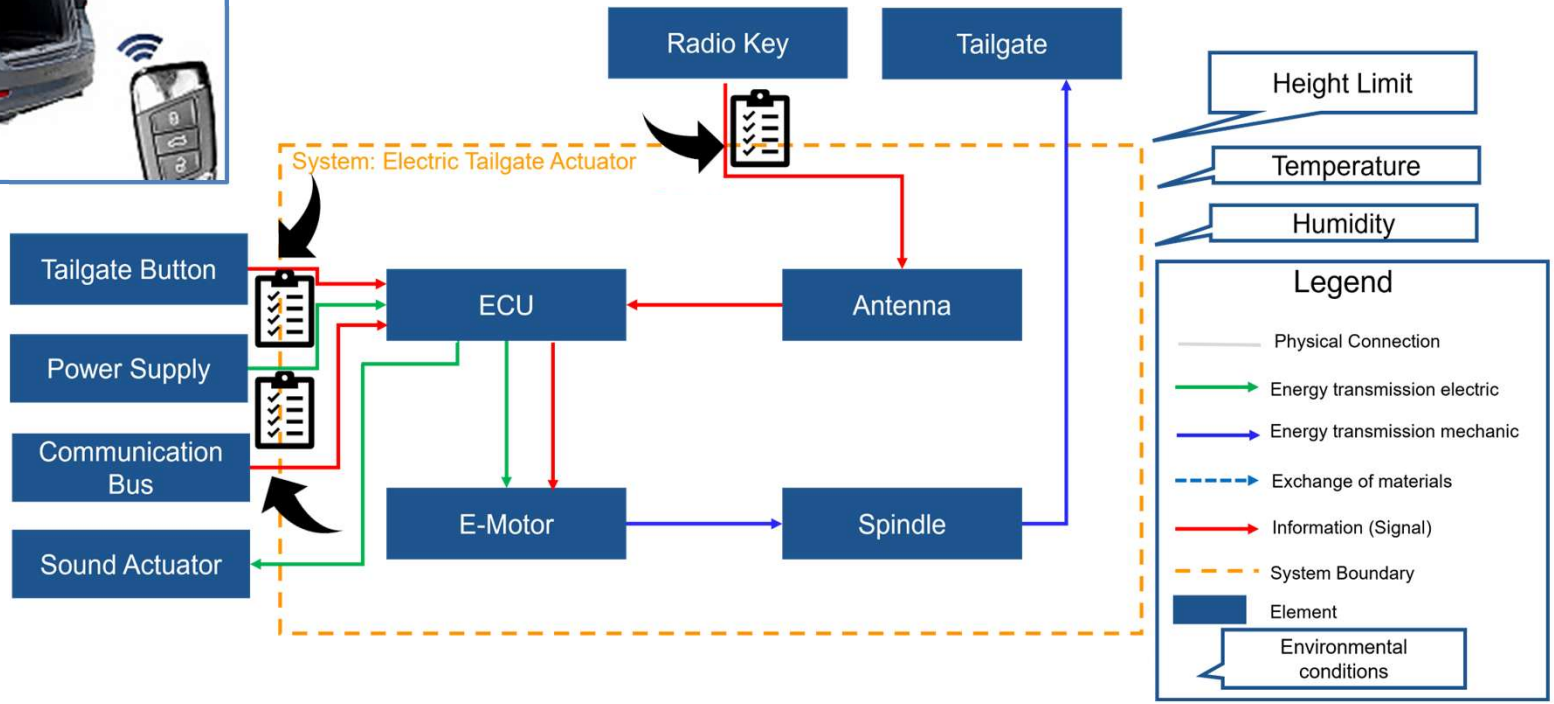
Understand Where AI Creates Value — and Where Trust Must Be Earned

Compare Human and AI-Assisted Creation: Assess the effort, quality, completeness and project usefulness of AI-generated engineering work products compared to traditional approaches.

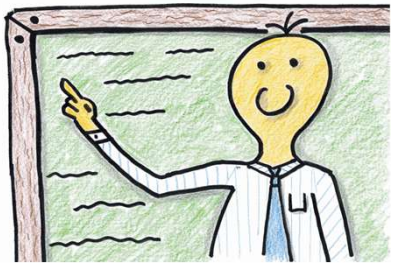
Benchmark Review Effectiveness: Evaluate how humans and AI differ in detecting defects, inconsistencies and traceability gaps in engineering artifacts.

Distinguish Real Evidence from "Assessment-Ready" Artifacts: Develop practical criteria to assess provenance, authenticity, traceability and actual project usage of AI-supported evidence.

Workshop Example: Electric Tailgate Actuator System



Exercise 1: Checking AI Work Products: "AI vs Human Defect Hunt (1/2)



<< Review of the created System Requirements >>

After creation of the system requirements of our tailgate system:

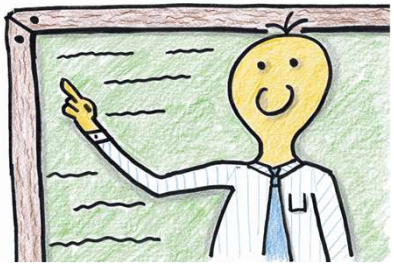
Each Group:

- **First part (20 min):**
  - Manually review a defined set of requirements

- **Second part (20 min):**
  - Review a defined set of system requirements with an GenAI tool
  - Investigate different tools and compare the results
  - Use different prompting strategies

Time: 40 min.

Exercise 1: Checking AI Work Products: "AI vs Human Defect Hunt (2/2)



<< Review of the created System Requirements >>

Result presentation:

First part:

- How many findings have you detected (manually)? What was your approach?

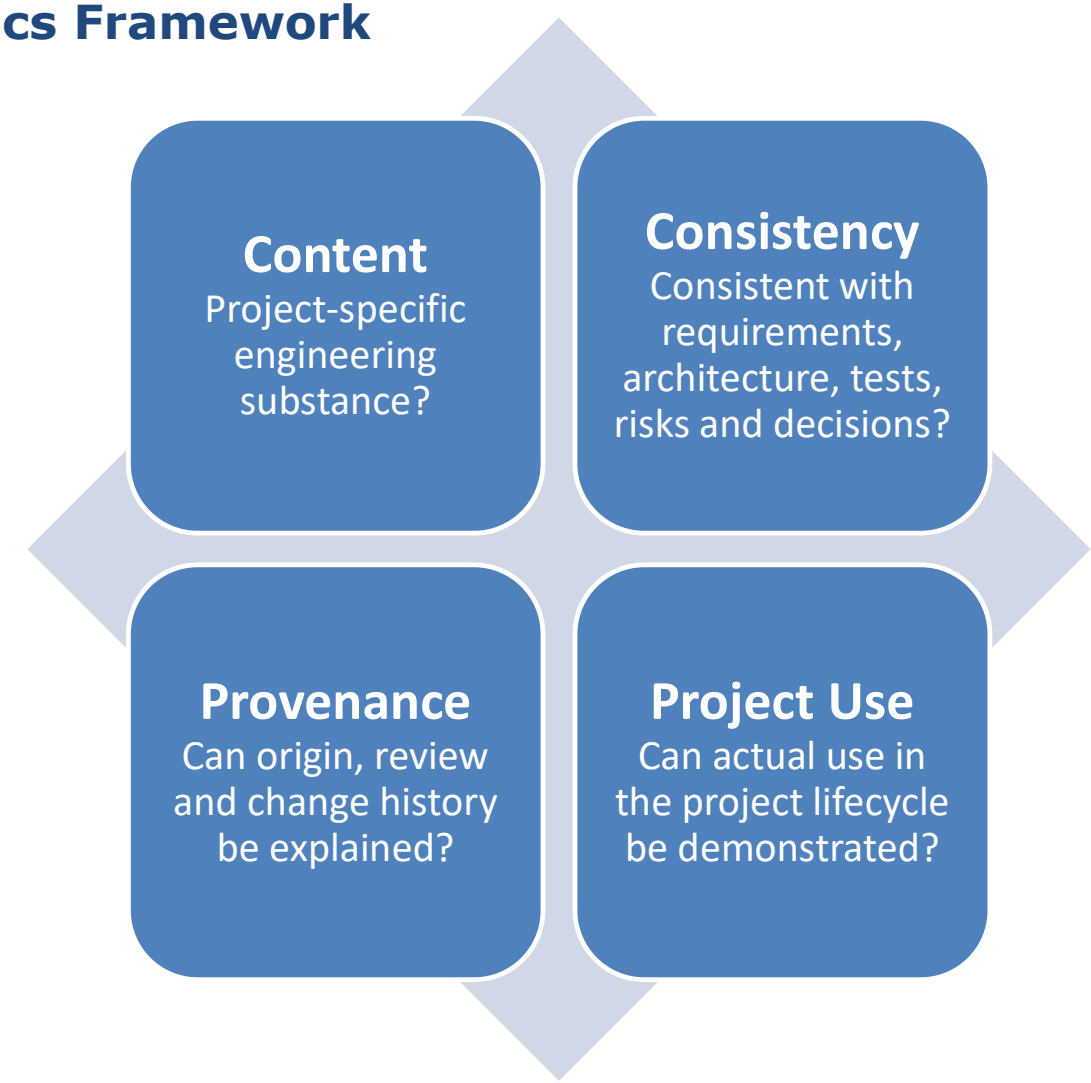
Second part:

- How many findings have you detected? Which tool delivered the best results?
- How many different prompts did you try? What was the most promising one?

Time: 20 min.

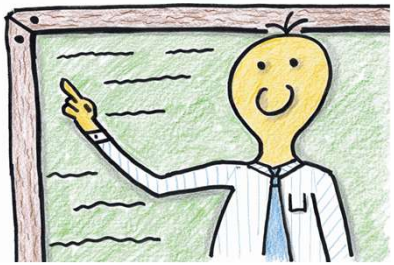
A Simple Forensics Framework

Fake evidence is rarely detected by one single indicator.
Use **four dimensions** to structure your assessment.



Do not only inspect the document — **follow the traces** around it.

Exercise 2: Recognizing Fake Evidence: "Audit Forensics"



<< Detection of fake evidence >>

Imagine you are in an assessment scenario and perform the interview for several ASPICE processes. The ASPICE assessment uses both manually generated work products as well as AI generated ones.
Your job is now to question the AI generated WPs if they were purely generated for the assessment as "Fake evidences" or actively in use within the project.

Please discuss the following points in both groups:

- What are your experiences with AI generated WPs?
- In which processes do they occur the most?
- How can you detect AI generated WPs? → **"Detection criteria"**

Time: 30 min group discussion + 30 min plenum discussion

Workshop Documentation

Exercise 1: Checking AI Work Products: "AI vs Human Defect Hunt

Setup:

- 4 groups were formed with different experience levels
- Each group needed to manually review and review it with the support of AI

Outcome manual review:

- Manual review was more time consuming, not every group managed to finish in time
- Good alignment within in the group on content discussion
- **Manual review identified better results in terms of quality**

Outcome AI review:

- All groups could finish the review within the time constraints
- Used tools: ChatGPT, Copilot, Claude, Gemini -> **In this setting, Claude was perceived as strongest by the participants.**
- AI struggled with the traceability detection between stakeholder and system requirements
- AI added up content which was not specifically asked for
- Most of the groups used a role-based prompting approach

Workshop Documentation

Exercise 2: Recognizing Fake Evidence: "Audit Forensics"

Outcome of the discussion:

- **What are your experiences with AI generated WPs?**
  - PMP, Test Reports, Templates, Checklists, Technical Documentation
- **In which processes do they occur the most?**
  - SWE.1, SYS.2, MAN.3, MAN.5, SUP.1, SWE.3, SWE.4, SWE.6, SYS.4, SUP.10
- How can you detect AI generated WPs? → **"Detection criteria"**
  - **Timestamp, review logs, linguistic perfection, inconsistency with practical application, ownership, semantic limitation on traceability, perfect polished wording, too much text, too generic results**

- **All "fake" evidences have been detected by the groups**

Key Takeaways

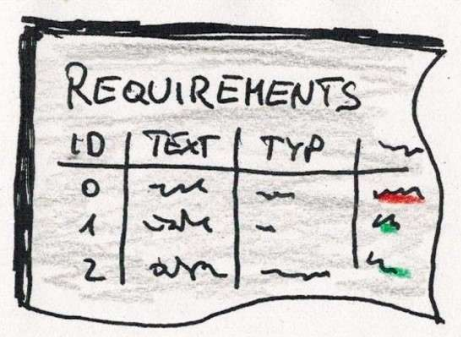
- AI supported review should be a **preliminary review at first**, **human check needed**
- **Human review provide a better quality** but is of course more time consuming
- Critical number of requirements should be checked, when is it worth to detect with AI
- **Results by AI are sometimes nebulous because "unasked" content is added** (e.g. misleading categorization)
- AI detection criteria should be established or more explicitly checked (e.g. t**imestamp, review logs, linguistic perfection, etc.)**

Creating artifacts is no longer the bottleneck **-> Validation, integrity, and demonstrating real project use have become the new challenge!**

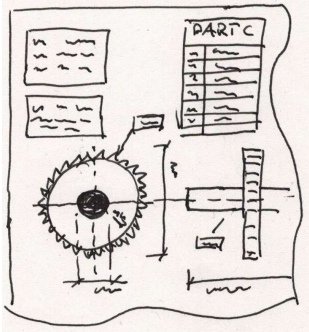
Fake evidence is not detected by style alone — it is detected by missing project substance.

From Workshop Observations to Practice

What we focus on in the next part:

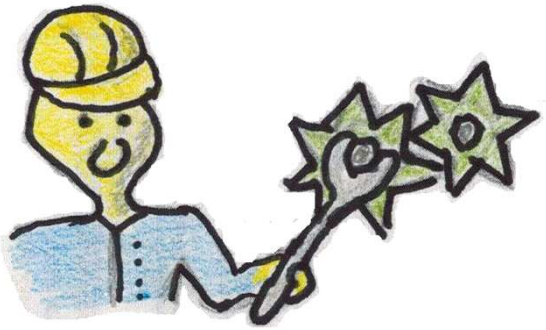


(Review-) Criteria



(Engineering) Context

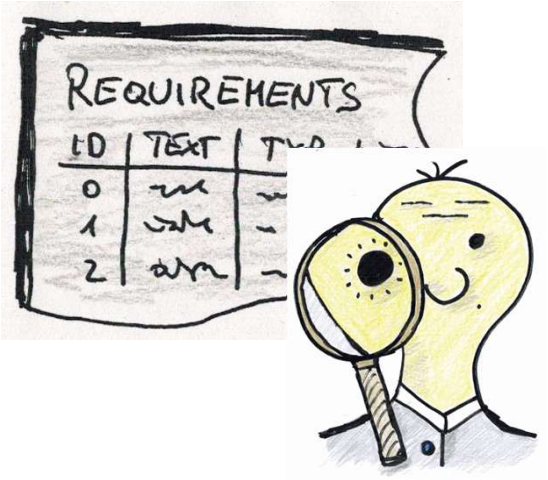
Responsibility



Toolchain

Human Expertise

(Review) Criteria: AI Needs a Strategy



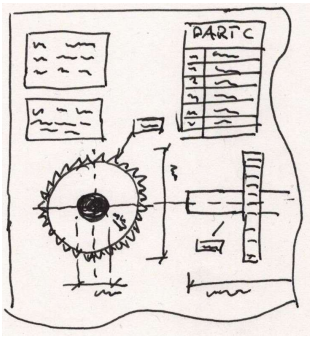
AI does not automatically know what **"good"** means!

Better prompts help — but explicit **criteria** matter more.

In the workshop, the strongest findings came from **systematic checks**.

Traceability & Coverage	Check if every requirement is linked.	Check whether every system requirement is linked to at least one stakeholder requirement, whether the link is meaningful, and whether all stakeholder requirements are sufficiently covered by system requirements.
Consistency of Values and Constraints	Check if the requirements are consistent.	Compare all values, units, tolerances, timing constraints, operating conditions and temperature ranges between stakeholder and system requirements. Flag every deviation that is not explicitly justified.

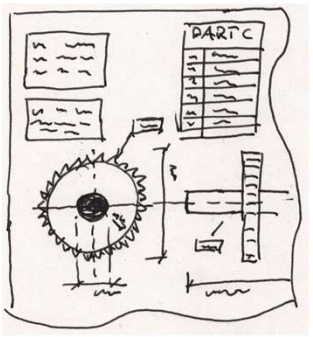
Engineering Context: AI Needs the Whole Picture



	Without context	With context
Stakeholder Requirements	accept plausible wording	check whether values, tolerances and constraints are preserved
Architecture & Interfaces	miss design relevance or interface gaps	check consistency with system structure and responsibilities
Variants & Configuration	overlook invalid assumptions	verify whether requirements apply to the right product variant
Test Evidence	accept “passed” as sufficient	ask for measurement data, limits, logs and result rationale
Change & Review History	treat polished text as mature evidence	check whether decisions, findings and changes are traceable

Without engineering context, AI reviews text.
With engineering context, AI can support engineering review.

Engineering Context: AI Needs the Whole Picture



	Without context	With context
Stakeholder Requirements	accept plausible wording	check whether valu… and co…
Architecture & Interfaces	miss dep…	consisten…
…ge & Review History	treat polished text as mature evidence	check whether decisions, findings and changes are traceable

ID	System Requirement
SYS-TG-011	The tailgate actuator system shall support powered opening and closing operation in the ambient temperature range from -40 °C to +85 °C.

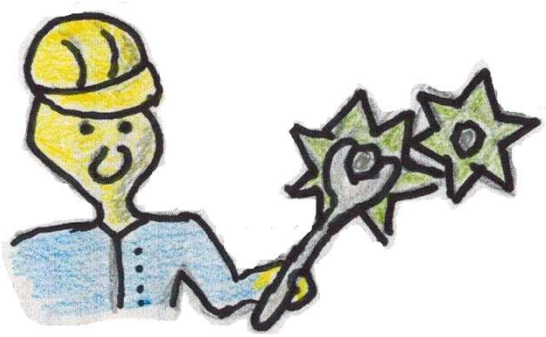
STRS.8	Non-Functional	The system must be able to be used in the temperature range from -30 °C to +55 °C“.

…ation … apply … duct variant

ask for measurement data, limits, logs and result rationale

Without engineering context, AI reviews text.
With engineering context, AI can support engineering review.

Engineering Toolchain: AI Must Fit Your Way of Working



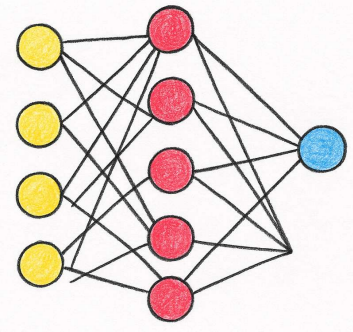
Typical engineering sources:

- Requirements management
- Architecture and interface descriptions
- Test management
- Defect and issue tracking
- Review records
- Change requests and baselines
- Variant and configuration management

Is AI working with isolated document snippets — or with the actual engineering environment?

AI must **support the engineering workflow** — not create a parallel documentation world.

AI Toolchain: AI Needs Access, Boundaries and Control



- Which AI tool is suitable for the task?
- Which data may the AI access?
- Does the AI have access to the relevant project context?
- Are requirements, tests, reviews and defects connected?
- Are prompts, outputs and decisions documented?
- Who reviews and approves AI-supported results?
- Are confidentiality, IP and compliance constraints addressed?

Without the right AI setup, even a powerful model remains a risky shortcut.

Human Expertise: AI Still Needs Judgement

- define the right task and review criteria
- provide domain and project context
- challenge plausible but wrong results
- detect inconsistencies that “look reasonable”
- decide whether deviations are acceptable
- approve and take responsibility for the outcome

AI can generate answers.
Engineers must ask the right questions.

Human Expertise: AI Still Needs Judgement

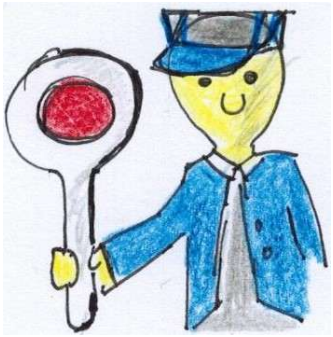
- define the right task and review criteria
- provide domain and project context
- challenge plausible but wro
- detect inconsi
- d

ID	System Requirement
SYS-TG-011	The tailgate actuator system shall support powered opening and closing operation in the ambient temperature range from -40 °C to +85 °C.

STRS.8	Non-Functional	The system must be able to be used in the temperature range from -30 °C to +55 °C“.

generate answers.
Engineers must ask the right questions.

Responsibility: AI Can Support — But It Cannot Own the Evidence



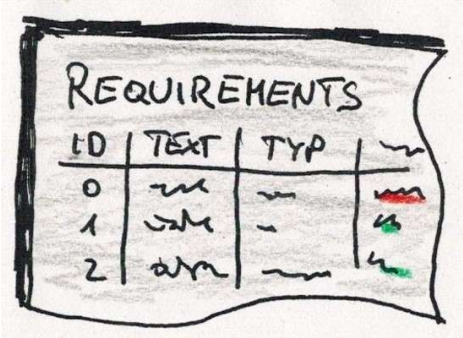
The project team must be able to explain:

- why the work product exists
- which input was used
- what AI contributed
- how the result was reviewed
- where it is used in the project
- which human decision was made based on the AI-supported proposal

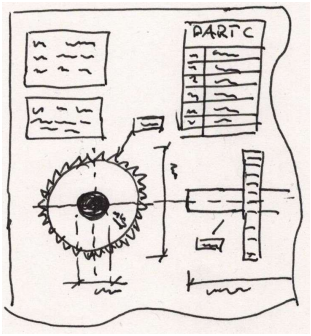
Using AI does not remove engineering responsibility.

It can provide suggestions, drafts and review findings.
But **ethical and engineering responsibility** requires a human decision.

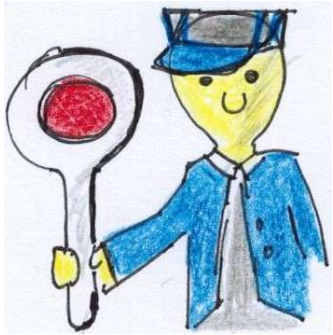
Conclusion: From AI Output to Trustworthy Evidence



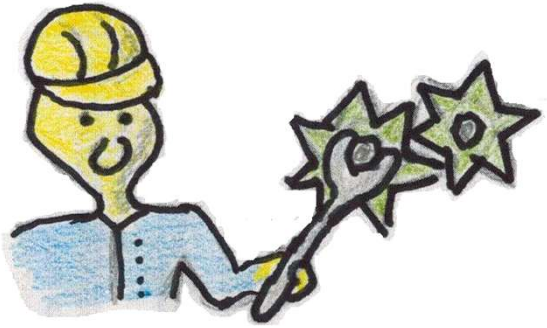
(Review-) Criteria



(Engineering) Context



Responsibility



Toolchain

Human Expertise

The future is not **human vs. AI**.
The future is **human-led engineering** with AI-supported evidence.

Discussion: Would You Trust AI?

- Where do you already see AI-supported engineering work products?
- Which work products would you allow AI to create or review?
- What would you need before trusting AI-supported evidence?
- Which part is most critical in your organization: criteria, context, toolchain, expertise or responsibility?

Let’s discuss.

Process Fellows GmbH | Schlegelleithe 8 | 91320 Ebermannstadt | GERMANY

Phone: +49 9194 3719 957

Website: www.processfellows.de | E-Mail: info@processfellows.de