Responsible Engineering of Information Systems Based on Generative Artificial Intelligence

An Action Design Research Study at a German Premium Car Manufacturer

Dominik Fetzer, Henner Gimpel, Oliver Meindl, Jan Strickmann

Business & Information Systems Engineering (2025)

Appendix (available online via http://link.springer.com)

Appendix A – Interview Details of the ADR Project at Premium Automotive Group (PAG)

Appendix A.1 – Involved Practitioners

Practitioner	Position	Experience in AI (in years)	Continuous Support	Contribution in Stage		
				Diagnosis	Design	Implementation
P1	IT architect	10	X	X	X	X
P2	IT security manager	2	X	X	X	X
P3	IT project manager	8	X	X	X	X
P4	Project manager from HR department	4	X	X	X	X
P5	Program manager	7	X	X	X	
P6	GenAI engineer	2	X	X	X	X
P7	IT quality manager	2		X	X	
P8	Product owner	5		X	X	
P9	GenAI engineer	8			X	X
P10	IT quality manager	1		X	X	
P11	IT architect	5			X	
P12	IT compliance manager	2			X	
P13	In-house legal counsel	10		X	X	
P14	IT project manager	3				X
P15*	Use case owner	None				X
P16*	Use case owner	None				X
P17*	Data owner	None				X
P18	IT project manager	1				X
P19	GenAI engineer	3				X

*Included due to strong domain knowledge (e.g., HR knowledge that the HR assistant should advise on)

The selection of practitioners for the diagnosis and design stages was conducted solely by the research team. All practitioners were recruited from PAG's GenAI initiative, which brought together individuals with extensive AI experience and advanced GenAI knowledge. Key selection criteria for the core practitioners (P1 - P6) included their central role in PAG's GenAI initiative due to their skills and the coverage of a broad range of competencies relevant to the responsible engineering of GenAI-based IS. Additional practitioners were selected role-specific and purposeful, some also recommended by the eADR core team, to incorporate skills and experiences vital for method development that had not yet been covered. In the implementation stage, organizational factors influenced the selection. For example, the inclusion of certain practitioners was essential due to their roles and the specific domain knowledge they possess (e.g., P15, P16, P17). Additionally, PAG selected some practitioners to contribute to the engineering of the three use cases based on their competencies.

Appendix A.2 – Overview of the Conducted Interviews

Interview #	Practitioner	eADR-Stage	Purpose of the Interview	Duration (in Minutes)	Evaluation Criteria	Illustrative Interview Information
1	P1	Diagnosis**	Approaching the research problem from a practical perspective	between 20 and 30 each	• Problem importance (Sonnenberg and vom Brocke 2012)	Interviewee-specific discussions
2	P10					
3	P2					
4	P3					
5	P5					
6	P6					
7	P4					
8	P8					
9	P9					
10	P7					
11	P1	Diagnosis**	Evaluation of the problem definition, the project's goal, and the meta-requirements	32	• Understandability of problem and meta-requirements (Sonnenberg and vom Brocke 2012) • Completeness of meta-requirements (Sonnenberg and vom Brocke 2012; Tuunanen et al. 2024) • Suitability of project goal (Sonnenberg and vom Brocke 2012)	Appendixes A.3 – A.4
12	P2			56		
13	P3			43		
14	P4			46		
15	P5			44		
16	P6			38		
17*	P1, P2, P3, P4, P6	Design – Cycle #1**	Evaluation of the initial method draft with a focus on MeRGE's overarching structure (i.e., the rough division of the activities into different phases)	89	• Level of detail of the phases (Sonnenberg and vom Brocke 2012) • Completeness of the activities (Sonnenberg and vom Brocke 2012) • Consistency of the phases (Sonnenberg and vom Brocke 2012; Tuunanen et al. 2024) • Clarity of structure (Sonnenberg and vom Brocke 2012)	Appendixes A.5 – A.6
18	P6			31		
19	P2			59		

20	P6	Design – Cycle #2**	Evaluation of MeRGE's fine structure within the individual phases with a focus on the inputs and outputs, activities, roles, and tools	17	• Understandability of activities (Sonnenberg and vom Brocke 2012) • Conciseness of activities (own) • Consistency of activities, roles, and tools (Sonnenberg and vom Brocke 2012; Tuunanen et al. 2024) • Fidelity of method with real world phenomenon (Sonnenberg and vom Brocke 2012)	Appendixes A.7 – A.8
21	P5			28		
22	P3			35		
23	P2			16		
24	P1			15		
25	P4			37		
26	P2			21		
27	P5			33		
28	P6		Evaluation of MeRGE regarding its practicality across multiple dimensions and the achievement of the project goal as well as the fulfillment of the meta-requirements	26	• Fit of the method to the validated problem statement and project goal (Tuunanen et al. 2024) • Feasibility of the method (Sonnenberg and vom Brocke 2012; Tuunanen et al. 2024) • Utility of the method application (Tuunanen et al. 2024) • Efficiency of the method's utilization (Sonnenberg and vom Brocke 2012; Tuunanen et al. 2024)	Appendixes A.9 – A.10
29	P3			51		
30	P1			27		
31	P4			29		
32	P2			37		
33	P7		Evaluation of MeRGE with further practitioners not part of the ADR core team to broaden the evaluation perspective	31	• Completeness of all method's components per stakeholder requirements (Tuunanen et al. 2024) • (Anticipated) efficacy of the method (Tuunanen et al. 2024) • (Anticipated) utility of the method application (Tuunanen et al. 2024)	Appendixes A.11 – A.12
34	P8			75		
35	P9			66		
36	P10			32		
37	P11			29		
38	P12			58		
39	P13			53		

40*	P4, P6, P14, P15, P16, P17, P18	Implementation HR Assistant**	Goal-oriented evaluation of applying MeRGE in situ based on the realization of three use cases	64	• Operationality (Sonnenberg and vom Brocke 2012) • (Realized) utility of the method application (Tuunanen et al. 2024) • Impact on artifact environment and user (Sonnenberg and vom Brocke 2012) • *Further evaluated through closely accompanied implementation and supplementary interviews*	Goal: Establishing a basic understanding of the use case
41*	P15, P16			30		Goal: Expanding the understanding of the use case
42	P16			29		Goal: Preparation of the validation data set
43*	P15, P16			77		Goal: Presentation of an interim version of the HR Assistant
44	P4			43		
45	P14			71		Goal: Use case-specific evaluation of the method application Appendixes A.13 – A.14
46	P18	Implementation PAGGPT		78	• Applicability of method to use cases (Sonnenberg and vom Brocke 2012) • Ex-post completeness of method per stakeholder requirements (Tuunanen et al. 2024) • Projectability (Tuunanen et al. 2024) • Generalization (Sonnenberg and vom Brocke 2012; Tuunanen et al. 2024)	
47	P9			87		
48	P19	Implementation PM Assistant		141		

*Focus group interview
**Conducted without reference to the three different use cases

Appendix A.3 – Interview Guide #1 "Diagnosis Stage"

Content Input: We have presented our current understanding of the problem definition.

1. To what extent do you agree with the problem definition and the derived findings?
2. In your opinion, which key findings were forgotten or which aspect was neglected in the description of the problem?

Content input: We have presented our current definition of the eADR project's goal, including the environment for which the method is intended, the initial state before, and the target state after the application of the method (i.e., the situation-specific requirements for the application of the intended method).

1. What do you think of the eADR project's goal? Do you think it needs to be narrowed down or expanded?
2. What are your thoughts on the defined application context? Please also include possible useful extensions or restrictions.

Content input: We have presented the five overarching requirements for the method, i.e., the meta-requirements that we have derived.

1. How do you assess these overarching requirements, particularly regarding their understandability?
2. What requirements for a future solution do you miss?

Appendix A.4 – Aggregated Outcomes: Interview Guide #1 "Diagnosis Stage"

Problem definition:

All interviewed practitioners confirmed the presented problem definition, affirming its clarity and relevance.

Two practitioners (P3, IT project manager, and P6, GenAI Engineer) with experience across multiple organizations beyond PAG emphasized the broader significance of the presented problem. They highlighted that the challenges described are not unique to PAG but are also faced by other organizations. Moreover, the lack of widespread integration of GenAI in organizations highlights the need for comprehensive and structured guidance on responsibly engineering GenAI-based IS. These two practitioners further noted that they have been approached for such guidance multiple times, reinforcing the importance of addressing this gap.

EADR project's goal:

All practitioners concurred with the researcher's initial definition of the ADR project's goal, describing it as well-formulated and thoroughly comprehensive, leading to no significant revisions. However, two experts noted that the goal is quite extensive, highlighting the necessity to strike a balance between detail and comprehensiveness. Additionally, it was suggested that the method should ideally also serve as a checklist, covering all relevant areas to be considered.

Situation-specific requirements for the application of the intended method:

Regarding the situation-specific requirements for applying the method, the evaluation findings resulted in the most significant adaptations during the diagnosis stage, particularly concerning the intended environment for the method. The expert evaluation highlighted making the following constraints:

- Confining the method to using pre-trained GenAI models as a service through APIs.
- Excluding the training, development of GenAI models from scratch, and hosting.
- Excluding the additional fine-tuning of GenAI models provided as a service.

Concerning the *initial state* before and the *target state* after the application of the method, there was unanimous agreement among the experts. The GenAI use case to be realized constitutes the method's starting point. Thus, the method is not intended as a "*brainstorming*" tool for identifying GenAI use cases within organizations. The experts described this limitation as sensible at the current time. Post-application, the target state is the GenAI use case being realized and introduced within the organization. The experts concurred that it is appropriate to exclude the long-term operation of GenAI-based IS from consideration. Addressing this topic should constitute the content of a separate research project and a future extension of the method, given the broad scope of long-term operations.

Meta-requirements:

The experts evaluated the meta-requirements as understandable, sensible, and comprehensive, particularly highlighting their broad spectrum positively. Three experts proposed an additional meta-requirement. Two of these suggestions were identified by the researchers as concrete expressions of existing meta-requirements and were thus considered to be already addressed. The third proposed meta-requirement was categorized as overly generic and applicable to any method. Therefore, the evaluation did not lead to any changes regarding the meta-requirements.

Appendix A.5 – Interview Guide #2 "Design – Cycle #1"

The foundation for the focus group interview and the two follow-up interviews was the initial method draft presented in Appendix F. This draft was provided to the interview participants via a Miro board, allowing collaborative interaction during the interview (the Miro board was used during all interviews as a central tool for visualizing the current state of the method). At the start of the interview, the entire method was covered, with sections revealed progressively as the interview proceeded.

Content Input: Structuring the method into five phases – Uncovering the phase names and providing a brief three-sentence description for each phase.

1. What are your thoughts on the division into five phases, and what do you think of the individual phases?
2. Are there any phases you feel are missing, or do you consider any phases to be unnecessary?

Content Input: Activities – Uncovering and brief presentation of the Precheck phase activities.

1. Task for practitioners:
   We would like your feedback on the individual activities. Please complete this task using post-its directly on the Miro board. Mark all activities as follows:
   - Green: retain & correctly located.
   - Yellow: generally good, but requires slight modification or is incorrectly located.
   - Red: remove/superfluous.
   - Blue: Which steps should be particularly emphasized or have special importance in engineering GenAI-based IS?
2. Discussion of activities that were not marked exclusively in green to explore the reasons behind these evaluations.
3. Which activities do you find missing in this phase?
4. Repetition of the process for the other four phases.

Content Input: Longitudinal activities – Uncovering and a brief presentation of the longitudinal activities

1. What do you think of the overarching longitudinal activities?
2. Do you miss any longitudinal activity?

Final Question:

1. Is there anything else you have noticed in general that you would like to suggest for revising the method?

Appendix A.6 – Aggregated Outcomes: Interview Guide #2 "Design – Cycle #1"

Structuring the method into five phases:

- The five phases were evaluated as well-organized, understandable, sensible, and comprehensive.
- No phase was deemed superfluous or missing.
- The significant importance of the 'Precheck phase' and the classification of data with regard to its confidentiality was highlighted multiple times ("*We often rush too quickly ahead.*" – P4, Project manager from HR department).
- In the second and third phases, the "*what*" and the "*how*" of realization are mixed; the 'Conceptualization phase' (formerly 'Initiation phase') should focus on the "*what,*" while the 'Development phase' should address the "*how.*"
- The necessary iterative approach, especially in the middle three phases, is not yet sufficiently clear: "*We should be careful with the entire approach to ensure that it does not appear too much like a waterfall.*" (P1, IT architect).

Activities:

- Precheck:
    - The activity 'Risk analysis and risk-benefit assessment' can and should be conducted comprehensively only after the product vision is defined. Yet, it should already be performed to a reduced extent during the 'Precheck phase.'
- Conceptualization:
    - Activities related to the "*how*" of implementation should be shifted to the 'Development phase' (e.g., 'Preliminary evaluation of alternative GenAI models to determine which models could be considered').
    - Certain activities, such as 'Development of a realization strategy,' suggest a waterfall approach due to their rigid formulation and should be expressed more flexibly to align with agile development principles.
    - A market analysis to determine whether a readily available solution could be purchased is missing.
- Development:
    - The level of detail among activities varies significantly; certain activities are much more detailed (e.g., 'Implementation of an input control') compared to other activities within the 'Development phase' (e.g., 'Customized development of the GenAI use case') and those in other phases. Yet, the method should definitely include the detailed, GenAI-specific activities and treat them in depth.
    - The sequence of activities should be revised.

    - Certain activities are missing, such as the creation of a rough draft of the IS, which includes the development of a user interface and/or an integration into existing systems.
- Quality:
    - The activities ‘Checking the model security and model release for processed data’ and ‘Checking for possible copyright infringements with regard to the model output’ should occur a phase earlier.
    - The activity ‘Carrying out extensive and detailed testing of the use case’ should be deepened, as it poses a particular challenge in GenAI-based IS.
- Deployment:
    - The aspect of familiarizing employees with GenAI technology should be implemented centrally across the company rather than for each use case. Nonetheless, it is beneficial that this activity is included in the method.

Longitudinal activities:

The experts already mentioned the two longitudinal activities in response to our questions about missing activities following the discussion of the deployment phase before we uncovered them.

Appendix A.7 – Interview Guide #3: "Design – Cycle #2"

For this interview sequence, we present a non-exhaustive list of questions we asked, as they were partly very individual.

Understandability of activities, conciseness of activities:

- Which activities are not clearly formulated enough for you?
- Which activities would you rephrase?
- You were not satisfied with the formulation of the activity 'Restriction of the models' permissions;' how would you improve it?
- The activities 'Carrying out basic IT project protection' and 'Checking model security and model release for processed data' are very similar. How do you evaluate these two activities? Should these activities be consolidated?
- […]

Consistency of activities, roles, and tools:

- How do you feel about the restructuring of activities in the 'Development phase' and the introduction of sub-activities for 'Customized development of the GenAI use case?'
- What would you do differently compared to what is currently prescribed by the method?
- Which phase transitions do you consider necessary, and between which phases do you see particularly iterative back-and-forth switching?
- Which roles and required skills do you think are missing?
- Which tools and good practices that you would apply in individual activities in practice should be added? Which do you consider redundant?
- How would you test the output quality of a GenAI-based IS?
- […]

Appendix A.8 – Aggregated Outcomes: Interview Guide #3 "Design – Cycle #2"

Understandability of activities, conciseness of activities:

- The practitioners saw the potential for improvement regarding the clarity of the activities, for example:
  - The naming 'Handling of user prompts by the organization' is ambiguous because it is unclear which organization is being referred to.
  - "*We could rename the Initiation Phase 'Conceptualization;' this makes it sound less like it's only done once, and more like it's repeated in an agile approach.*" (P1, IT architect).
- The practitioners recommended consolidating some activities, for example:
  - The activities 'Selection of a GenAI model as a service provider' and 'Selecting appropriate GenAI model(s)' should be merged due to their close relationship.
  - The activity 'Checking model security and model release for processed data' was viewed as part of 'Carrying out basic IT project protection.' Furthermore, the wording of the latter activity should be improved, as it is unclear what is meant by "*model security.*"

Consistency of activities:

- Regarding the fine structure within the individual phases, the introduction of sub-activities for the activity 'Customized development of the GenAI use case' was evaluated as a very good solution.
- The focus of the iterations between phases is in the middle three phases, which should be emphasized even more.
- 'Testing the use case in the models' playground is considered a useful tool, not an activity.

Consistency of roles:

- Accurately describing certain roles was challenging (e.g., GenAI engineer and requirements engineer with broad GenAI understanding), as GenAI emerges new roles and skills.
- Some roles need specification. For example, the role of 'GenAI expert' is not specific enough and should be clarified as 'GenAI engineer' and 'requirements engineer with in-depth GenAI knowledge:' "*I don't think the role 'GenAI expert' is well chosen. 'GenAI expert' can mean anything – it's too vague as a term; the role needs to be defined more concretely.*" (P3, IT project manager).
- Certain roles were either entirely (e.g., 'IT architect') or partially missing in certain phases (e.g., 'Data owner' in the 'Quality phase').

- Some roles were considered optional, such as involving a ‘Data protection expert’ in the ‘Precheck phase.’

Consistency of tools:

- Despite the general consistency of the illustrative tools & good practices, additional suggestions were made for inclusion in the method (e.g., installing a smaller, upstream GenAI model to check whether user input is toxic, using disclaimers to exclude liability for inappropriate model outputs, or phased deployment).
- As some activities need further research (e.g., the ‘Testing’ activity in the ‘Quality phase’), the practitioners do not yet know an ideal solution for every activity.

Appendix A.9 – Interview Guide #4 "Design – Cycle #2"

Content Input: Presentation of the current state of the method on the Miro Board.

Content Input: Recap of the evaluated project goal.

1. To what extent do you see the project goal fulfilled by the method?
2. What advantages do you perceive as resulting from the method?
3. What challenges do you foresee associated with the method?
4. If required:
    - Discussion of meta-requirements that have not yet been addressed.
    - How do you assess the fulfillment of the meta-requirements by the developed method?

5. Which use cases do you find feasible with this method, and which are less so?
6. To what extent do you believe the method enhances the efficiency of engineering GenAI use cases? For whom might the method result in unjustified additional effort?
7. For which target groups do you see added value in the efficient engineering of GenAI use cases?
8. Final question: Is there anything else regarding the method that you would like to share with me?

Appendix A.10 – Aggregated Outcomes: Interview Guide #4 "Design – Cycle #2"

Fit of the method to the validated problem statement and project goal:

- The practitioners evaluated the project goal and meta-requirements as fulfilled: "*I consider the project goal to be completely achieved.*" (P2, IT security manager); "*I would proceed exactly as outlined in the method.*" (P6, GenAI engineer).
- One minor remark regarding the general applicability and completeness of the method was noted and should be examined in the future: as with nearly any method, it cannot be completely ruled out that edge cases exist where specific, highly domain-specific tasks need to be performed beyond the activities of MeRGE, or where additional, specialized roles might be required.

Feasibility of the method:

- The practitioners consider all use cases identified by PAG so far (approximately 80) to be feasible with the method.
- The method's flexibility also accommodates realizing use cases that do not only involve engineering type 1 ('Building of a dedicated GenAI-based application system') but instead focus on the integration into or extension of existing systems.
- Practitioners advise initially testing the method with smaller use cases and then tackling the "*tough nuts.*" (P1, IT architect).

Utility of the method application:

- The method provides project participants on both the IT and business department sides with a structured approach, offering significant value and a previously unavailable guideline: "*The great advantage is having something to guide you, ensuring that nothing is overlooked.*" (P3, IT project manager). In this vein, the method's big picture (i.e., Appendix G including the sub-activities elaborated on in Table 7) also serves as a checklist: "*For me, the method is also an excellent checklist to verify 'have we thought of everything?' […]. This is extremely helpful*." (P4, Project manager from HR department).
- Following the method ensures that no significant risks arise from introducing GenAI-based IS in organizations.
- The presentation of the method using a central big picture and the supplementary text was described as a very good and clear format. Yet, extensive descriptions of some activities would be useful in the next step, although this was not the focus and envisioned as part of the ADR project.

Efficiency of the method's utilization:

- *The method's utilization reduces the need for time and workforce, leading to efficiency gains, particularly:*
    - The method prevents avoidable loops during the engineering process: "*Just being able to orientate yourself on something saves you many loops that you might otherwise have to do if you don't quite know how it works.*" (P4, Project manager from HR department).
    - Adhering to the prescriptive method reduces the likelihood of things going bad or being overlooked from the outset, which need laborious and resource-intensive corrections at a later stage: "*If you follow the method properly, then you definitely save resources because you don't have to iron out things that went wrong from the start. That's always what eats up time and energy endlessly later on. It can become really costly and tough. By following the approach, you avoid nasty taskforce, hypercare, or whatever phases that become necessary when something has not worked or needs to be corrected*" (P2, IT security manager).

Final Question:

The practitioners had no further remarks. Instead, feedback such as "*We have now discussed the method very thoroughly. […] We have covered everything I wanted to address*" (P1, IT architect) was repeatedly given.

Appendix A.11 – Interview Guide #5 "Design – Cycle #2"

For this interview sequence, we present a non-exhaustive list of questions we asked, as they were partly very individual.

Completeness of all method's components per stakeholder requirements:

- From your perspective, what is essential for responsible GenAI-based IS engineering? What elements are included, and what could be improved?
- Can you identify with the method?
- What comments do you have about the method from a compliance perspective?
- In which phase do you believe an IT architect should be involved, and for which activities?
- Which activities need to incorporate legal aspects? What specific aspects and laws should these be?
- Do you have any further comments on the method?
- […]

(Anticipated) efficacy of the method, (anticipated) utility of the method application:

- What is your general opinion on such a technology-specific method?
- How do you view such a technology-specific method from your position as XXX?
- Does this method provide added value, and if so, for whom?
- From your perspective, is such a technology-specific method helpful for GenAI-based IS engineering projects?
- Does the method offer added value regarding the adequate representation of the legal perspective in GenAI-based IS engineering projects?
- In which areas have there been problems in the past that the method could address?
- […]

Appendix A.12 – Aggregated Outcomes: Interview Guide #5 "Design – Cycle #2"

Completeness of all method's components per stakeholder requirements:

- Apart from minor improvement suggestions, the method was considered complete at the chosen level of detail.
- Minor improvement suggestions were made regarding presentation, phrasing, and involved roles, such as:
  - "*The visualization seems very phase-oriented. The agile approach should be emphasized more*." (P7, IT quality manager).
  - The activity 'Carrying out basic IT project protection' was not precise enough, making its exact meaning unclear.
  - The phrasing "*Establishing human oversight*" (P12, IT compliance manager) clarifies the activity's significance more effectively than 'Integrating a human loop'. The activity should be renamed accordingly.
  - In the 'Precheck phase,' an 'AI compliance manager' suffices if required, as they can quickly identify which use cases should not be pursued further. 'IT security managers' and 'Legal experts' are not yet required at this stage.
- Additionally, the content of several activities was discussed in detail regarding the method's textual description.

(Anticipated) efficacy of the method, (Anticipated) utility of the method application:

- The method is generally assessed as highly useful and important, e.g., "*I think a specific method is a very good idea because the subject matter is so complicated that you can no longer simply say, well, just do scrum.*" (P8, Product owner).
- Anticipated advantages of MeRGE include:
  - Efficiency benefits in terms of speed and resource savings.
  - The method increases transparency and awareness among all stakeholders, including management, regarding the complexity and specific requirements associated with the introduction of GenAI: "*We can also use this method to sensitize our colleagues in the business departments to the fact that it's not that simple. Yes, because otherwise – and this is also a very old disease – IT, which deals with the topic, is perceived as the brakeman. If you make the method transparent, then people will see what's behind it.*" (P11, IT architect); "*The management always makes life easy for itself and says, it's actually the same as before, just a little bit extra, so I don't need to get any extra budgets, and then those in technology have to pay for it, and that's why I find the method very charming.*" (P8, Product owner).
  - In addition to providing a coordinating effect, which was lacking in previous technology hypes, the introduction of the method is also attributed a regulatory effect: "*So I also*

see it as an aid to provide the company itself with a conscience protection against itself. […]. People are so driven by hypes and chase after hypes that they completely neglect essential security aspects, ethical aspects, data protection aspects, and I don't know what other aspects, and this creates risks. These risks can be massively reduced or mitigated by such a method." (P11, IT architect).

Appendix A.13 – Interview Guide #6 “Implementation”

For this stage, we present a non-exhaustive list of questions we asked, as they were partly very individual.

Applicability of method to use cases:

- Do you have any immediate comments on the method application in your case that you would like to share directly?
- Was the method applicable to you in your use case?
- How did you engineer activity XXX in your use case?
- What could be presented in a more accessible manner?
- […]

Ex-post completeness of method per stakeholder requirements:

- To what extent is your approach reflected in the method?
- Did you take steps that are not represented by the method?
- What activities do you find missing in the method?
- […]

Projectability, generalization:

- Can you imagine engineering other, also larger use cases with the method?
- What limitations do you see with the method regarding its application in other use cases you know and in other organizations?
- Do you see any aspects that restrict the application of the method to certain use cases or to [PAG]?
- […]

Appendix A.14 – Aggregated Outcomes: Interview Guide #6 "Implementation"

Applicability of method to use cases:

The practitioners confirmed the applicability across all use cases, in particular also for the PR Assistant and PAGGPT:

- "*Yes, definitely. The method fully reflects our approach.*" (P19, GenAI engineer).
- "*The steps fit well and align with the steps we conducted in the project. I don't see anything missing at the moment.*" (P18, IT project manager).

Ex-post completeness of method per stakeholder requirements:

Aside from minor improvement suggestions regarding the extent of the textual description of the method, it was considered complete at the chosen level of detail:

- "*We followed the activities of the method and did not miss any steps.*" (P18, IT project manager).
- "*No, I didn't miss anything. [...] The only thing I would really go into more detail about is testing [...]. And I think an even more detailed text would actually be good.*" (P19, GenAI engineer).

Projectability, generalization:

Affirmation of the projectability and generalization of our method, even beyond the context of PAG, was given:

- A practitioner also involved in the engineering of GenAI use cases in organizations beyond PAG stated: "*I know over 60 use cases […]. All the use cases I'm aware of can be implemented with the method.*" (P9, GenAI engineer).
- "*In any case, you can implement other, even larger use cases with it. [...] I can't think of anything that would limit the method to specific use cases or to PAG.*" (P19, GenAI engineer).

Appendix B – Detailed Research Design Adapted From Mullarkey and Hevner 2019

	Diagnosis	Design #1	Design #2	Implementation
Mullarkey and Hevner (2019)	▪ Analysis of the significance of the problem to be solved for research and practice (p. 4) ▪ Definition of the goal of the eADR project (p. 4)	▪ Identification and conception of the proposed artifact design (p. 5) ▪ Creation of designs that incorporate innovative ideas to solve the given problems (p. 5) ▪ Collaborative intervention with co-creation activities (p. 5)		▪ Instantiation of artifacts through an implementation activity at the partner organization (p. 5) ▪ Evaluation of the proposed design (p. 5)
Main Activities	▪ Analysis of previous findings of PAG's GenAI initiative ▪ Carrying out a literature search to better understand and scientifically substantiate the problem ▪ Regular participation in internal company meetings and preparation of field notes ▪ Ten initial interviews for a problem understanding from a practical perspective ▪ Six additional evaluation interviews	▪ Identification of relevant internal methods ▪ Literature search to identify relevant methods from research ▪ Development of an initial method draft using the assembly-based SME approach ▪ Participation in internal meetings and creation of field notes ▪ Focus group discussion with five practitioners and two follow-up interviews ▪ Development of an initial version of the method	▪ Recategorization of some activities in the Development phase ▪ Addition of roles & required skills of participating stakeholders and illustrative tools & good practices ▪ Participation in internal meetings, creation of field notes, and close dialogue with practitioners ▪ 20 practitioner interviews ▪ Revision and final refinement of the method	▪ Application and evaluation of the method based on the realization of three GenAI use cases ▪ Three focus group discussions with seven practitioners as well as 6 additional interviews
Created Artifact	▪ Problem definition ▪ Definition of the goal of the eADR project ▪ Five meta-requirements derived from the problem and the defined goal ▪ Situation-specific requirements for applying the method	▪ Initial version of the method: • Five preliminary phases with matched activities • Particularly iterative exchange between the middle three phases • Two longitudinal activities	▪ Final version of the method: • Five phases • Detailed activities • Roles & required skills • Illustrative tools & good practices	▪ Fully realized and launched HR Assistant ▪ Prototypically realized PM Assistant ▪ Fully realized and launched PAGGPT
Evaluation Outcomes	▪ The problem has been fully understood and exists in research and practice ▪ A holistic guideline for the responsible engineering of GenAI-based IS is currently lacking but urgently needed ▪ Additions are necessary concerning the situation-specific requirements for applying the method	▪ The five phases were evaluated to be coherent ▪ Compared to the initial method draft, some activities need to be reorganized into other phases, split up, modified, or supplemented ▪ The activities in the Development phase vary significantly in level of detail ▪ The exchange between phases should be adapted	▪ The elements of tools and roles must be modified ▪ Strong iterative relationship between the middle three phases was emphasized again ▪ Method fulfills the initially defined goal and meta-requirements	▪ The method was perceived as easy to use and very helpful for realizing the three use cases ▪ The three use cases could be realized efficiently and effectively using the method ▪ Some features of the engineered GenAI-based IS the method prescribed were described as very useful by the users
Reflection & Learning Outcomes	▪ Development of a holistic methodological guideline for the responsible engineering of GenAI-based IS in organizations was set as the goal of the eADR project ▪ A suitable balance must be struck between the comprehensiveness and depth of detail of the method	▪ The goal to develop the overarching structure of the method was achieved ▪ The upcoming ADR cycle should concentrate on GenAI-specific activities ▪ The evaluations with practitioners were perceived as very valuable; to strengthen the ADR principles of mutually influencing roles and authentic and concurrent evaluation, these should be intensified	▪ The goal to add roles and skills, optimize the fine structure, and establish conciseness was achieved ▪ Various ADR principles (e.g., reciprocal shaping, mutually influential roles, authentic and concurrent evaluation) were better accounted for through closely timed exchanges with practitioners ▪ Our method takes into account the ADR principle of generalized outcomes	▪ PAG has successfully completed all phases of the method at least once, implying that the practitioners can apply the method in future GenAI integrations ▪ The method must be tested further in practice realizing further use cases of varying complexity, size, and modality and expanded with further useful illustrative tools & good practices, if necessary

Appendix C – Meta-Requirements for the MeRGE Method

Meta-requirements create a generally valid solution that is not tailored to a specific use case or organization. The derivation and use of meta-requirements is a standard procedure in the context of design-oriented research (Gregor and Hevner 2013) and is also explicitly mentioned by Sein et al. (2011) in the context of the ADR research paradigm. They can be used to evaluate and assess the developed artifact regarding whether it fulfills the defined meta-requirements (Pries-Heje et al. 2008). Although other requirements may also be of interest for the engineering of GenAI-based IS in organizations, we consider the following five to be the most relevant, especially with regard to GenAI's specific challenges and the responsibility areas.

- *MR1*: Reduction of familiarization time in GenAI-based IS engineering projects for all employees involved
- *MR2*: Reduction of the project risk of GenAI-based IS engineering projects for the organization
- *MR3*: Improving the quality of GenAI-based IS engineering projects while saving resources for the organization
- *MR4*: Increasing the benefits that the utilization of GenAI offers end users
- *MR5*: Improving collaboration between the various stakeholders involved in GenAI-based IS engineering projects

Appendix D – Creation of the Initial Method Draft Applying the Assembly-Based SME Approach

To create our initial method draft, we followed the assembly-based SME approach. This approach entails three steps: 1) *determining the situation-specific requirements for using the method*, 2) *selecting suitable method chunks*, 3) *assembling the selected chunks into a new, situation-specific method* (Ralyté et al. 2003). We have already defined the situation-specific requirements in the diagnosis stage and will discuss these in more detail in section 4.1. To select suitable method chunks, that is, theoretical elements that we can inscribe actively into our method following the ADR principle of the *theory-ingrained artifact*, existing methods must be identified and broken down into chunks. These must be stored in a so-called method base (Ralyté et al. 2003). We drew on a variety of sources to create our method base. These included methods that already exist internally at PAG as well as methods identified in a literature search relating to the engineering of IS, software, and AI. As the existing methods do not adequately reflect the specific challenges and requirements of utilizing GenAI, we also relied heavily on previously published scientific papers dealing with individual aspects of the responsible engineering of GenAI-based IS (e.g., Weidinger et al. 2021; Kasneci et al. 2023; Zhao et al. 2023). In addition, the findings of PAG's GenAI initiative and from the initial interviews were incorporated. The result was a method base with 44 method chunks (see Appendix E). We rechecked all chunks for suitability and compliance with the specified method requirements (Ralyté et al. 2003), excluding four chunks. In the final step, we assembled the selected method chunks to create a new, situation-specific method that meets the requirements of the specific situation and, if necessary, adapted them to the situational needs (Ralyté et al. 2003). Since the work system life cycle proposed by Alter (2001) is basically valid for our goal; we used its phases (initiation, development, implementation, and we renamed the latter deployment) for the initial structuring of our new method and assigned the chunks. In addition to minor adjustments such as combining chunks, we introduced two additional phases, Precheck and Quality, as we have identified some activities upstream of the Initiation phase and activities explicitly related to control and quality assurance.

Appendix E – Method Base (Condensed Version)

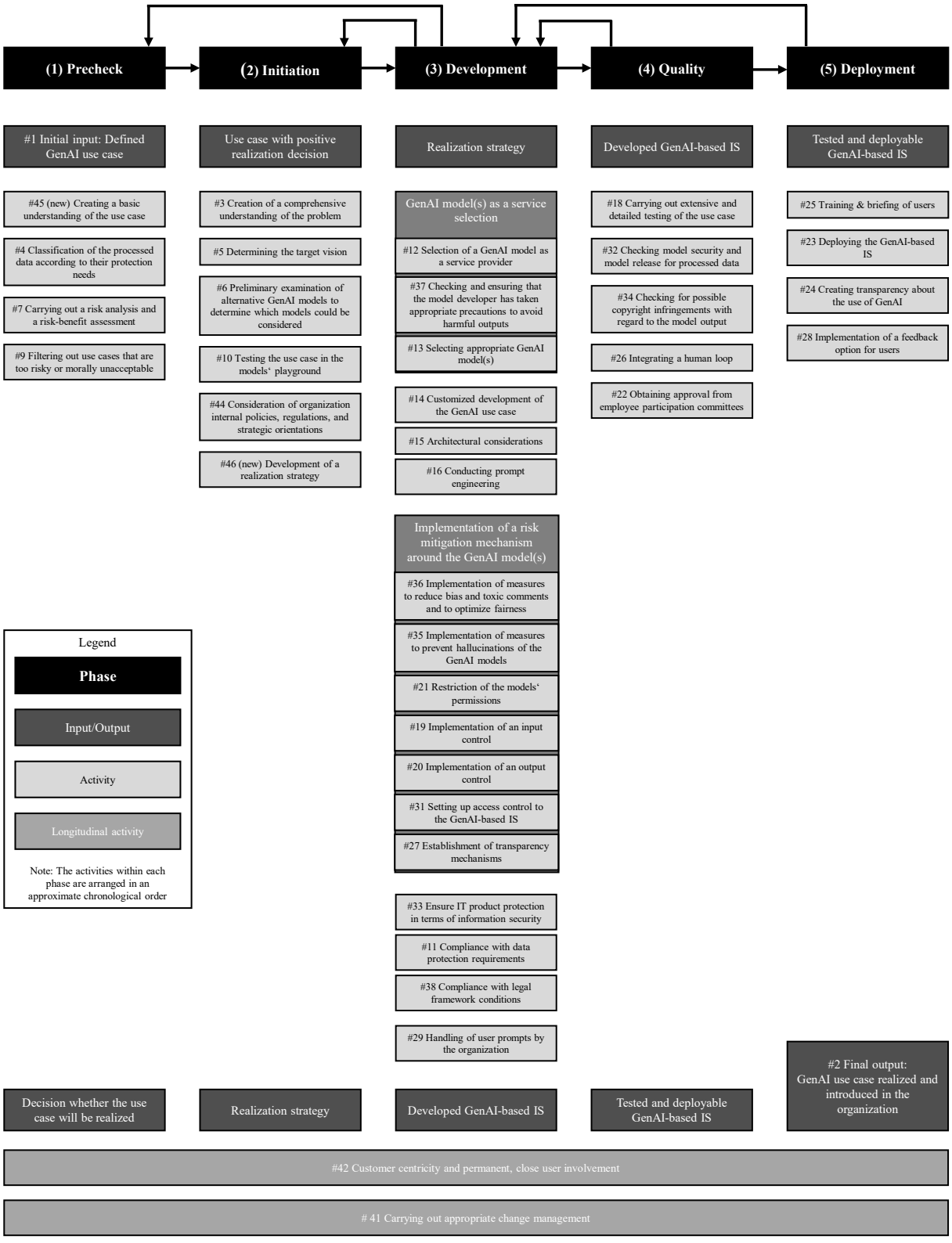
Chunk ID	Short Description	PAG Internal Source	Illustrative Literature Sources (Existing Methods | GenAI-Specific Sources)	Roles	Tools & Techniques	Included in the Initial Method Draft
#1	Initial input: Defined GenAI use case	✓				✓
#2	Final output: GenAI use case realized and introduced in the organization	✓				✓
#3	Creation of a comprehensive understanding of the problem	✓	Alter 2001; Radack 2009; Leau et al. 2012 |	Use case owner	GenAI use case canvas	✓
#4	Classification of the processed data according to their protection needs	✓	Radack 2009 |	Use case owner, Data owner, Data protection expert	Organization internal data classification framework	✓
#5	Determining the target vision	✓	Alter 2001 |	Use case owner, End user, GenAI expert	Customer journey	✓
#6	Preliminary examination of alternative GenAI models to determine which models could be considered	✓	| Hötter and Warmuth 2023	GenAI expert		✓
#7	Carrying out a risk analysis and a risk-benefit assessment	✓	| Weidinger et al. 2021; Haleem et al. 2022	GenAI expert, IT security manager, Data protection expert, Use case owner	Risk analysis	✓
#8	Classification of the use case according to the risk class of the European AI Act	✓	| Weidinger et al. 2021; Haleem et al. 2022; Ray 2023	Legal expert, Use case owner, End user	European AI Act (prohibits or severely restricts certain use cases)	✓ (Included in other activities)
#9	Filtering out use cases that are too risky or morally unacceptable	✓		GenAI expert, IT security manager, Data protection expert, Use case owner	Internal and general ethical guidelines, European AI Act	✓
#10	Testing the use case in the models' playground	✓		GenAI expert, IT project manager	Playground of the models	✓
#11	Compliance with data protection requirements	✓	| Kasneci et al. 2023; Ray 2023; Khowaja et al. 2024	IT security manager	GDPR/data protection laws	✓
#12	Selection of a GenAI model as a service provider	✓	Alter 2001 | Kucharavy et al. 2023; Mökander et al. 2023; OWASP 2023	GenAI engineer	Tools/frameworks for comparing GenAI models, Benchmarks, Model information	✓

#13	Selecting appropriate GenAI model(s)	✓	Alter 2001 | Hötter and Warmuth 2023	GenAI Engineer	Tools/frameworks for comparing GenAI models, Benchmarks, Model information, Playground of the models	✓
#14	Customized development of the GenAI use case	✓	Alter 2001; Leau et al. 2012 | Hötter and Warmuth 2023	GenAI engineer, Software developer	PoC, MVP, prototyping, agile software development methods, source code management tools, GenAI-specific frameworks (e.g., LangChain)	✓
#15	Architectural considerations	✓	Leau et al. 2012 | Dwivedi et al. 2023; Zhao et al. 2023	GenAI engineer		✓
#16	Conducting prompt engineering	✓	| Hötter and Warmuth 2023; Zhao et al. 2023; Hariri 2024	GenAI engineer, Use case owner, End user	Prompt patterns, Zero-, One-, Few-shot-prompting, In-context learning, Prompt-chaining, Chain-of-thought prompting	✓
#17	Fine-tuning the GenAI model	✓	| Kasneci et al. 2023; Khowaja et al. 2024	GenAI engineer		Outside the method's scope
#18	Carrying out extensive and detailed testing of the use case	✓	Leau et al. 2012; Spjuth et al. 2021 | Mökander et al. 2023; Hariri 2024	GenAI engineer, Software developer, Use case owner, End user, IT project manager	Functional tests, Task-specific tests, Usability tests, Load and performance tests, Test management software	✓
#19	Implementation of an input control		| OWASP 2023			✓
#20	Implementation of an output control		| Weidinger et al. 2021; OWASP 2023; Yan 2024		Implementation of a small, downstream model that checks whether the model output is manipulated or ethically incorrect, Syntactic validation	✓
#21	Restriction of the models' permissions		| OWASP 2023			✓
#22	Obtaining approval from employee participation committees	✓		IT project manager, Works council		✓
#23	Deploying the GenAI-based IS	✓	Alter 2001; Radack 2009; Spjuth et al. 2021 |	IT project manager, GenAI engineer, Software developer	Piloting, Soft launch, Creation of a communication concept	✓
#24	Creating transparency about the use of GenAI		| Hötter and Warmuth 2023; Ray 2023; Hariri 2024	Software developer	Prominent display of a notice on the user interface	✓

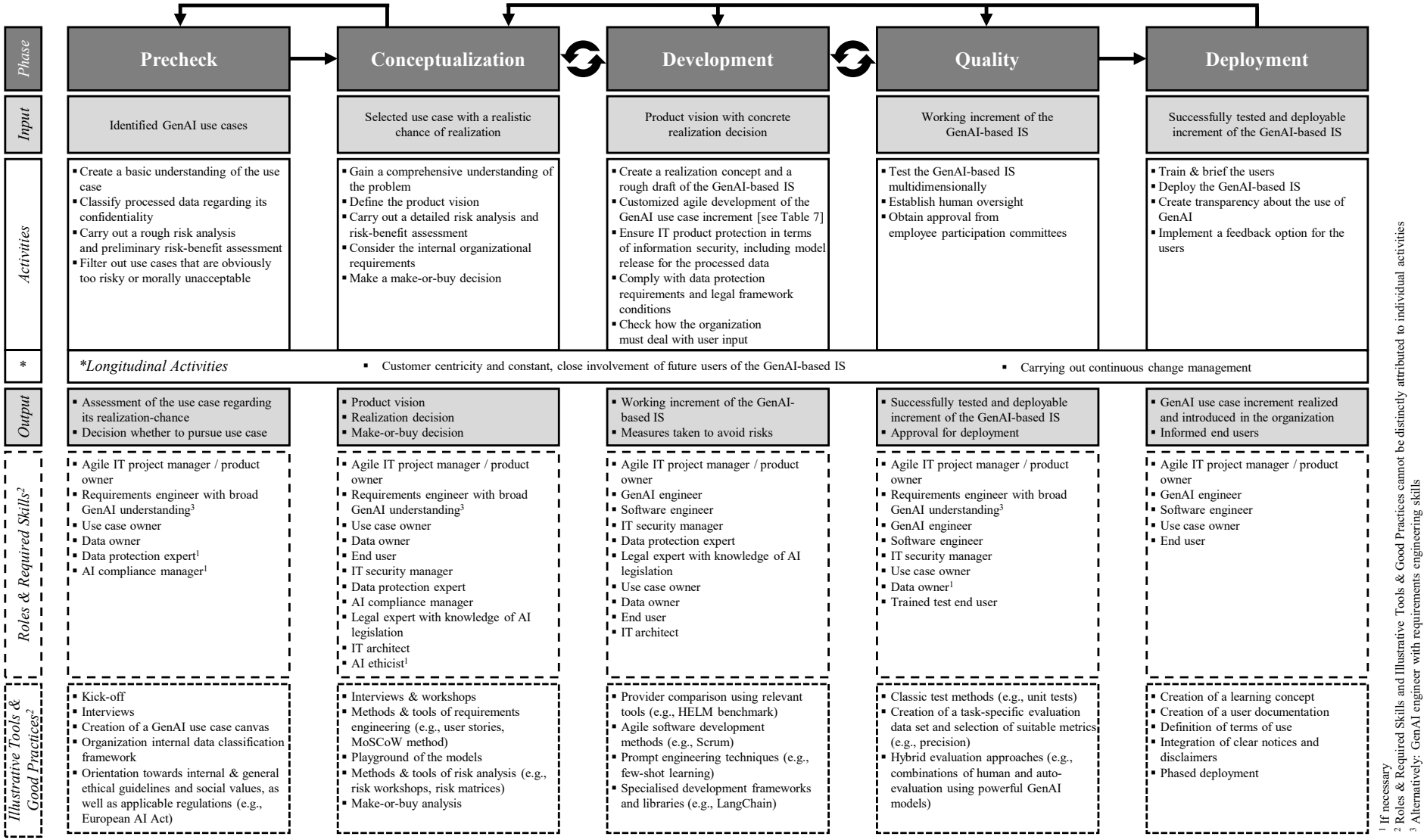
#25	Training & briefing of users	✓	Alter 2001 | Dwivedi et al. 2023; Kasneci et al. 2023	AI project manager	Learning concept, User documentation, Terms of use, Clear notices, Disclaimers	✓
#26	Integrating a human loop	✓	| Haleem et al. 2022; Kasneci et al. 2023; Khowaja et al. 2024		Implementation of a user approval	✓
#27	Establishment of transparency mechanisms	✓	Suresh and Guttag 2021 | Kasneci et al. 2023; OWASP 2023	Software developer, GenAI engineer	Possibility for the user to check the content generated by GenAI	✓
#28	Implementation of a feedback option for users	✓	Leau et al. 2012 | Mökander et al. 2023; Khowaja et al. 2024		Thumbs up/down rating, Free text rating	✓
#29	Handling of user prompts by the organization	✓	| Ray 2023; Hariri 2024	Software developer, IT security manager		✓
#30	Ongoing monitoring of the GenAI-based IS	✓	Radack 2009 | Kasneci et al. 2023; Mökander et al. 2023		Regular review of the outputs of the GenAI-based IS	Outside the method's scope
#31	Setting up access control to the GenAI-based IS	✓	Spjuth et al. 2021| Ray 2023; Khowaja et al. 2024	IT security manager, Software developer	Multifactor authentication	✓
#32	Checking model security and model release for processed data	✓	| OWASP 2023	IT security manager	Internal IT security standards, Organization internal data classification	✓
#33	Ensure IT product protection in terms of information security	✓	Radack 2009 | Ray 2023; Hariri 2024	IT project manager, IT security manager, Software developer, GenAI engineer	Internal IT security standards, Penetration test, Vulnerability scan	✓
#34	Checking for possible copyright infringements with regard to the model output	✓	| OWASP 2023	IT project manager, Legal expert	Copyright law	✓
#35	Implementation of measures to prevent hallucinations of the GenAI models	✓	| Kaddour et al. 2023; Peng et al. 2023; Asgharzadeh et al. 2023	GenAI engineer	Content moderation components	✓
#36	Implementation of measures to reduce bias and toxic comments and to optimize fairness	✓	| Weidinger et al. 2021; Hötter and Warmuth 2023	GenAI engineer	Implementation of a small, downstream model that doublechecks the model output	✓
#37	Checking and ensuring that the model developer has taken appropriate precautions to		| Weidinger et al. 2021; OWASP 2023			✓

	avoid harmful outputs					
#38	Compliance with legal framework conditions	✓	Bonazzi et al. 2009 | Block et al. 2023; Mökander et al. 2023	IT security manager, Data protection expert, Legal expert	European AI Act, Data Act, Copyright law	✓
#39	Adhere to internal and external ethical guidelines	✓	| Ray 2023; Zhuo et al. 2023; Khowaja et al. 2024	AI ethicist	General and internal ethical guidelines, General social values, Codes of conduct	✓ (Included in other activities)
#40	Documentation of the project	✓		IT project manager	Documentation software (e.g., Confluence), Documentation templates	chunk too generic
#41	Carrying out appropriate change management	✓	Stowasser and Suchy 2022 |	IT project manager, Change expert	Tools of change management, Development of a change concept	✓
#42	Customer centricity and permanent, close user involvement	✓	Leau et al. 2012 |	IT project manager		✓
#43	Implementation of risk management	✓	| Gimpel et al. 2020	IT project manager	Risk management system	chunk too generic
#44	Consideration of organization internal policies, regulations, and strategic orientations	✓		IT project manager	Organization internal policies, guidelines, regulations and strategic directions	✓

Appendix F – Initial Method Draft



Appendix G – Detailed Overview of the MeRGE Method



Appendix H – Application of the RAG Pattern in Use Case "HR Assistant"

In engineering the HR Assistant, we applied the RAG pattern to prevent hallucinations and ensure that the HR Assistant answers questions solely based on the HR knowledge available on the intranet. The RAG pattern is a technique designed to retrieve targeted information from external knowledge sources (e.g., databases) to generate informed and contextually relevant answers (Gao et al. 2023). Vector databases are often used as external knowledge sources because they provide several functionalities required by RAG pattern, such as semantic search (Gao et al. 2023). Therefore, we also utilized a pgvector database for the HR Assistant. HR knowledge from selected intranet pages is extracted via API at regular intervals, particularly when the intranet pages are updated. Then, the extracted HR knowledge is divided into smaller segments of 1,000 characters each, also called chunks, enabling targeted information retrieval. Each chunk is annotated with source information (i.e., the corresponding intranet page), and for each chunk, using an embedding model, a vector embedding is generated. A vector embedding is a numerical representation of the chunk in the form of a multi-dimensional vector, which represents the semantic meaning of the chunk. All chunks and their embeddings are stored in the pgvector database. When a user submits a question to the HR Assistant, a vector embedding is generated for the query as well. A hybrid search, combining keyword search with embedding-based similarity matching, is then conducted in the pgvector database, where similar data have similar embeddings. In this way, the chunks most relevant to the query are retrieved from the pgvector database. The GenAI model is provided with these retrieved chunks and finally generates an answer based exclusively on this knowledge. When the knowledge base encompasses a vast number of chunks across diverse topics, incorporating an additional step in the form of determining the topic area can be beneficial (e.g., regarding response quality). Subsequently, a search can be conducted exclusively within the chunks relating to the identified topic area, rather than searching the entire database.

Appendix I – Our Nascent IS Design Theory

This table summarizes our nascent design theory, describing the eight components proposed by Gregor and Jones (2007).

Component	Description
Purpose and scope	The aim is to develop a method that provides a prescriptive guideline for the responsible engineering of GenAI-based IS in medium and large-sized organizations.
Constructs	GenAI-based IS, responsible engineering, organizations, generative variability, AI models as ready-to-use components, (MeRGE) method, high-level architecture
Principle of form and function	We provide a five-phase method enriched with roles & required skills, illustrative tools & good practices, enabling medium and large-sized organizations to exploit GenAI's potential while mitigating potential risks.
Artifact mutability	The designers consider the diversity of GenAI use cases using an agile approach that allows for flexible engineering of GenAI-based IS, satisfying the specific use case with its individual risks and each organization's risk tolerance.
Testable proposition	Applying this method to realize GenAI use cases leads organizations to mitigate unintended, negative consequences of utilizing GenAI.
Justificatory knowledge	Extant knowledge of GenAI, IS engineering methods, responsibility areas, and previous initiatives for responsible engineering of IS.
Principles of implementation	All practitioners realizing GenAI use cases should be familiar with the method to fully utilize its potential. Each organization should decide how the method is integrated into its processes, considering the existing policies and processes. The prescribed activities serve as guidelines and should be considered. However, the actual realization of the activities and the extent of their fulfillment must be decided on a case-by-case basis.
Expository instantiation	We instantiated the MeRGE method by realizing three GenAI use cases at PAG.

References

Sources included in Appendix A to Appendix I that go beyond the references of the paper:

Alter S (2001) Which life cycle: Work system, information system, or software? CAIS 7. https://doi.org/10.17705/1CAIS.00717

Bonazzi R, Hussami L, Pigneur Y (2010) Compliance management is becoming a major issue in IS design. In: D'Atri A, Saccà D (eds) Information systems: people, organizations, institutions, and technologies. pp 391–398. https://doi.org/10.1007/978-3-7908-2148-2_45

Gao Y, Xiong Y, Gao X, Jia K, Pan J, Bi Y, Dai Y, Sun J, Wang M, Wang H (2023) Retrieval-augmented generation for large language models: A survey. arXiv. https://doi.org/10.48550/arXiv.2312.10997

Stowasser S, Suchy O (2022) Künstliche Intelligenz erfolgreich einführen: Gestaltungsansätze für das Change-Management in Unternehmen. Digitale Welt 6:77–79. https://doi.org/10.1007/s42354-022-0564-4

Suresh H, Guttag J (2021) A framework for understanding sources of harm throughout the machine learning life cycle. In: Proceedings of the 1st ACM conference on equity and access in algorithms, mechanisms, and optimization. ACM, pp 1–9. https://doi.org/10.1145/3465416.3483305

Zhuo TY, Huang Y, Chen C, Xing Z (2023) Red teaming ChatGPT via jailbreaking: Bias, robustness, reliability and toxicity. arXiv. https://doi.org/10.48550/arXiv.2301.12867