

Operationalizing the EU AI Act in Agile Software Development: A Guideline-Based Approach

Dennis Schrader 

University of Applied Sciences and Arts Hannover, Germany

Eva-Maria Schön 

University of Applied Sciences Emden/Leer, Germany

Henning Fritzemeier 

Volkswagen AG, Wolfsburg, Germany

Michael Neumann 

University of Applied Sciences and Arts Hannover, Germany

Abstract

Context. The EU AI Act requires providers and deployers of Artificial Intelligence (AI) systems to implement documentation, risk management, and human oversight. Agile teams that ship AI features in short iterations lack specific artifacts to discharge these duties, since the regulation's abstract provisions do not map onto agile practices such as the Definition of Done, Sprint Reviews, or working agreements.

Objective. We provide agile teams with an actionable compliance instrument: an evaluated guideline that operationalizes EU AI Act obligations as activities integrable into existing agile practices. We further document the translation method to make its potential adaptation to other regulations examinable.

Method. Following Design Science Research, we assessed each EU AI Act article along three dimensions (*technical, normative, organizational*). We subsequently classified the articles using a traffic-light scheme and mapped those deemed highly relevant to previously documented pain points of agile teams working with AI. We evaluated the resulting catalog by eliciting practitioners' perceptions through a survey followed by semi-structured interviews with the same participants (N=11) and analyzed the data via qualitative content analysis.

Results. The guideline comprises 12 items covering roles and responsibilities, risk and quality management, transparency and traceability, monitoring, and regulatory sandboxes. Practitioners rated the catalog as understandable and relevant; feasibility varied with organizational maturity. Participants associated perceived adoption feasibility with collective ownership across roles and integration into existing agile events rather than parallel compliance processes.

Conclusions. The catalog gives agile teams a starting point to transform their delivery practices towards EU AI Act compliance without dismantling agile practices. A replication package is publicly available.

2012 ACM Subject Classification Software and its engineering; Human-centered computing → Empirical studies in collaborative and social computing

Keywords and phrases EU AI Act, AI Governance, Agile Software Development, Regulatory Compliance, Guidelines, Design Science Research

Digital Object Identifier 10.4230/LIPICs.ESEM.2026.94

Category Software Engineering in Practice Track Paper

Acknowledgements We thank all participants who participated in the survey and expert interviews for their time and valuable insights.



© Dennis Schrader, Eva-Maria Schön, Henning Fritzemeier, and Michael Neumann;
licensed under Creative Commons License CC-BY 4.0

20th International Symposium on Empirical Software Engineering and Measurement (ESEM 2026).

Editors: Robert Feldt, Maria Paasivaara, Daniel Mendez, Stefan Wagner, and Marvin Muñoz Barón;

Article No. 94; pp. 94:1–94:21



Leibniz International Proceedings in Informatics

Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Introduction

Organizations developing or integrating Artificial Intelligence (AI) components in agile or hybrid software development face a compliance problem that is now operational rather than prospective. The EU AI Act [6] entered into force in August 2024, and the core obligations for high-risk AI systems apply from 2 August 2026 [6, Art. 113]. Non-compliance carries administrative fines of up to 35 MEUR or 7 % of global annual turnover, whichever is higher [6, Art. 99]. Providers and deployers in the EU market therefore have to discharge lifecycle-wide duties for risk management, technical documentation, data governance, transparency, and human oversight [6, Art. 9, 11, 13, 14] on a fixed clock, while delivery teams continue to ship in two- or three-week Sprints.

Frameworks such as the *NIST AI Risk Management Framework* [24] and *ISO/IEC 42001* [12] primarily address organizational controls. A complementary gap remains at the team layer, where regulatory obligations must be translated into specific delivery activities and artifacts. For instance, in a specific aspect of the *Definition of Done*, *adoption in a Sprint Review*, *items in backlog refinement*, or *rows in a Working Agreement*. The EU AI Act is process-neutral by design: it prescribes outcomes, not ceremonies. A team applying a software process (e.g., Scrum, Kanban, or SAFe) receives no guidance on whether the *Art. 9* risk register lives in the backlog or beside it, whether *Art. 13* transparency obligations expand the user story template, or whether *Art. 14* human oversight is the Product Owner's, the team's, or a separate role's responsibility. Recent empirical work confirms that this gap is observable in practice: Neumann et al. [18] show in their study how regulatory pressures translate into policies that practitioners then bypass to maintain agile velocity, and Sillberg et al. [22] synthesize practitioner sentiment in a multivocal review and report widespread uncertainty about team-level implementation.

Similar research has produced compliance artifacts at the product, audit, or organizational level, which we discuss in Section 2. None of these contributions operates on the team layer, where the obligations meet the cadence of agile delivery, and the one approach that addresses an agile artifact is scoped to a single organization and lacks broader validation. What is missing is a team-level instrument that translates the obligations onto the agile events and artifacts an arbitrary team already uses, together with empirical evidence that practitioners find such an instrument understandable, relevant, and feasible to adopt.

Three research questions guide this work:

RQ1: *How can the requirements of the EU AI Act be systematically translated into actionable guidelines for the use of AI in agile software development?*

RQ1 aims to translate abstract regulatory text into team-level practice. We answer it by creating a reproducible procedure that assesses each EU AI Act article along three dimensions, classifies relevance through a traffic-light scheme, and maps high-relevance articles onto documented agile pain points [1, 8]. The expected outcome is both a method that is reusable for adjacent regulations and a concrete guideline catalog derived from it.

RQ2: *How do practitioners assess the guidelines regarding understandability, relevance, and feasibility?*

RQ2 verifies whether the artifact resulting from RQ1 holds up under practitioner scrutiny on three constructs that are diagnostic for an applied instrument: a) whether the guideline can be read without further interpretation, b) whether it captures the obligations that matter for agile teams, and c) whether it can be enacted in the carrier artifacts those agile teams already use. We expect this to surface design weaknesses and to indicate which clusters are robust across organizational contexts.

RQ3: *Which organizational factors do practitioners perceive as influencing the adoption of our EU AI Act guideline in agile teams?*

The last research question RQ3 moves from artifact quality to practitioners' perceptions of conditions that may support or hinder adoption. We probe how organizational size, regulatory exposure, agile maturity, and existing artifacts shape participants' assessments of potential adoption. The resulting conditions provide hypotheses for future field evaluation rather than evidence of actual adoption.

The paper makes three core contributions: First, we propose an approach for translating regulatory articles into specific (agile) practices. The procedure assesses each article of the EU AI Act along a technical, normative, and organizational dimension, classifies its relevance for agile teams through a traffic-light classification, maps known challenges while adopting or using agile methods from scientific literature, and aggregates them into thematic clusters. The approach is designed to be adaptable to regulations with a comparable obligation structure, such as DORA, NIS2, and the GDPR; however, this transfer has not been empirically evaluated and remains future work. Second, we present an EU AI Act adoption guideline consisting of 12 items organized into five clusters: roles and responsibilities, risk and quality management, transparency and traceability, monitoring and continuous compliance, and regulatory sandboxes. Third, we evaluated practitioners' perceptions of the guideline's understandability, relevance, and feasibility through a survey and semi-structured interviews with 11 practitioners; this study does not measure actual field adoption or compliance readiness.

The paper at hand is structured as follows: Section 2 positions our work against AI governance frameworks and related studies. Section 3 details the translation procedure and the evaluation design. In Section 4, we present the guideline catalog and in Section 5 the evaluation results. Section 6 discusses implications for practice, lessons learned, and the transfer of the methodology to other regulations. Section 7 addresses threats to validity before the paper closes with a conclusion and future work in Section 8.

2 Related Work

We position our work along the following research lines: the use of AI in agile software development and the operationalization of the EU AI Act for software engineering (SE). We close by stating the gap that our guidelines address.

2.1 AI in Agile Software Development

Empirical studies report that generative AI (GenAI) changes day-to-day agile work. Ulfsnes et al. [27] observe shifts in collaboration, pair work, and individual workflows once AI assistants enter the Sprint. Bahi et al. [1] review agile project management pain points that AI tooling may mitigate, from estimation support to backlog refinement. Russo et al. [21] argue in the Copenhagen Manifesto that GenAI in software engineering must remain human-centered, naming oversight, accountability, and developer agency as design constraints. Herda et al. [8] summarize the XP 2025 workshop in which practitioners and researchers map AI-and-agile challenges to a research roadmap.

Three recent contributions sharpen this picture. Planötscher et al. [20] survey 73 agile practitioners across seven software activities and report that AI is currently used as an assistant or not used at all, with practitioners expecting a shift toward AI as a collaborator. Perkusich et al. [19] survey 70 Brazilian practitioners on LLM use in Scrum management and find broad daily use alongside concerns about “almost correct” outputs, confidentiality, and hallucinations. Neumann et al. [18] conduct a multiple case study in three German

organizations and document how regulatory pressures translate into policies that practitioners then bypass to maintain agile velocity; one of their cases classifies GenAI tools by the EU AI Act risk pyramid, which is the most direct empirical evidence to date that the act is already shaping agile working practice. These studies establish that AI is now an agile concern and that the regulatory dimension is perceptible to practitioners. They do not, however, prescribe how teams should encode the resulting obligations in their daily practice.

2.2 Operationalizing the EU AI Act for Software Engineering

Several recent papers translate the EU AI Act into software engineering artifacts. Wagner and colleagues [28, 30, 29] introduce the regulation to the IEEE Software readership, dissect its requirements from an SE perspective, and add an industrial case study on high-risk readiness. Kelly et al. [13] extend the ISO/IEC 25059 product quality model with attributes derived from the act for safety-critical AI. Hupont et al. [11] contribute an UML-based use case card to document the intended purpose of a system in line with the risk classification, while Mökander et al. [16] discuss conformity assessment and post-market monitoring as auditing instruments. Hüger et al. [10] report from Cariad on an instantiation of the act through a Level of Done layer for ML projects in the automotive domain. Sovrano et al. [23] study how LLM-based tooling can support the drafting of technical documentation required by the act. Sillberg et al. [22] synthesize practitioner and academic sentiment toward the act in a multivocal review, and Westerstrand [31] argues that compliance does not exhaust ethical obligation and gives recommendations for fairness in agile AI development. Tuape et al. [26] address Article 14 *head-on and design mechanisms for human oversight and accountability* in AI-driven software engineering. Genovesi [7] describes the introduction of an AI governance framework in a financial organization as a practitioner report.

Each of these contributions advances a specific compliance artifact (a product quality model, a documentation template, an audit procedure, a Level of Detail (LoD) layer, a governance framework, a SE perspective on the regulation). None of them targets the team-level agile working practices through which providers and deployers actually meet their day-to-day obligations.

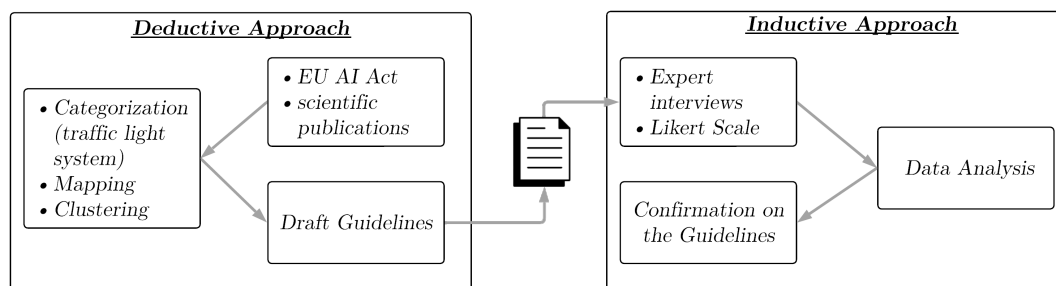
2.3 Research Gap

To the best of our knowledge, no published academic work translates the EU AI Act obligations onto the agile practices and artifacts an arbitrary team uses, or evaluates such a translation with practitioners across organizational contexts.

To be more precise, we identified three main gaps: First, the AI-and-agile literature recognizes that AI changes agile work and that the regulatory dimension is already perceptible in practice, with our own prior case study [18] documenting how teams improvise around the absence of clear rules; none of this work, however, supplies a team-level instrument that encodes the act's obligations. Second, the work on EU AI Act operationalization produces compliance artifacts at product, audit, or organizational level, and the one approach that explicitly addresses an agile artifact, Hüger et al.'s Level of Done [10], remains scoped to automotive ML and is not empirically validated with practitioners beyond the case organization. Third, agile-in-regulated-environments studies demonstrate feasibility for other regulations but precede the EU AI Act. This paper addresses these gaps.

3 Research Design

We designed our study by applying Design Science Research (DSR) [5, 9, 32] and organize it along the three cycles of Hevner [9]. DSR fits this work because the contribution is an artifact, a translation procedure and the resulting guideline, constructed and then evaluated through practitioner assessment, which is the configuration DSR was developed for in software engineering [5, 32]. The *relevance cycle* is grounded in a concrete industry problem: agile teams covered by the EU AI Act must satisfy obligations from 2 August 2026 onward [6, Art. 113] without abandoning iterative delivery. The *design cycle* produces a reproducible translation procedure and a guideline of 12 items (Section 4). The *rigor cycle* assesses both through a Likert survey and 11 semi-structured expert interviews analyzed through qualitative content analysis [15]. Figure 1 depicts our research design.



■ **Figure 1** Research Design for the development and evaluation of our guidelines.

3.1 Artifact Creation

The creation of the guideline proceeded in four phases (A–D). Each phase delivered an intermediate artifact for the next: a coded article set, a traffic-light classification, a mapping table, and the cluster structure.

Phase A: Deductive Analysis of the EU AI Act. We analyzed all articles of Regulation (EU) 2024/1689 along three dimensions:

- a *technical* dimension (data, monitoring, tooling),
- a *normative* dimension (obligations attached to provider or deployer roles under Art. 16 and 26),
- an *organizational* dimension (roles, governance, process integration).

Each dimension is anchored in recent empirical work on AI in agile contexts: technical in AI-assisted defect detection and test automation [25], normative in the alignment between AI Act obligations and ISO/IEC standards [13], organizational in roles, governance, and oversight in agile AI delivery [1, 8]. Coding criteria were derived deductively from the research questions and these references, following category-system construction as in Bortz and Döring [2] and the DSR guidance of Engström et al. [5].

Phase B: Traffic-light Classification. For each article and each dimension, we assigned *high*, *medium*, or *low* relevance. *High* means the article imposes obligations a team must address inside its development workflow; *medium* means the article is contextually relevant but generates no direct team-level work; *low* means it concerns governance levels beyond the team or actors other than providers or deployers. The aggregated rating per article is the maximum of the three dimensional ratings, since a single high-relevance dimension is

sufficient to make an article actionable. Articles with ambiguous coding were re-examined and, where uncertainty persisted, assigned the medium label so that ambiguity does not silently propagate. The first author performed the article coding and relevance classification; no independent sample was recoded and no inter-coder agreement was computed.

Phase C: Mapping to Agile Pain Points. We mapped every article rated high in at least one dimension to documented agile pain points of agile teams working with AI [1, 8], consolidated from the four studies cited above into recurring themes such as unclear AI ownership, lack of structured human oversight, insufficient data governance, and compliance documentation accumulating outside agile artifacts. The mapping is many-to-many and was refined iteratively to avoid uncovered pain points and redundant article coverage. We provide an example in Section 4.

Phase D: Cluster Building and Guideline Derivation. We grouped the mapped articles into thematic clusters following phases 1–5 of the type-building qualitative content analysis of Kuckartz and Rädiker [14]. Clustering produced five clusters (*C1 Roles and Responsibilities*, *C2 Risk and Quality Management*, *C3 Transparency and Traceability*, *C4 Monitoring and Continuous Compliance*, *C5 Regulatory Sandboxes and Innovation* under Art. 57). Within each cluster, we formulated one or more items through argumentative justification following Bortz and Döring [2], at a granularity that lets a team enact them within an existing event or artifact. Phase D did not use a separately enumerated pool of candidate items; therefore, no formally defined candidates were merged, removed, or rejected. Instead, the 12 items resulted directly from the cluster-wise derivation (C1: 2; C2: 3; C3: 2; C4: 2; C5: 3). We considered the set sufficient for the artifact’s intended scope because it covers every cluster and mapped problem–solution chain with a distinct team-level action. This establishes design coverage, whereas legal completeness remains an area for future work.

3.2 Practitioner Evaluation

The evaluation followed a sequential mixed-methods design: a Likert survey first, expert interviews second, so the interviews could probe the survey ratings. The same 11 practitioners were surveyed for both evaluation stages (see Table 1). Both data collection instruments are available in the research protocol, which we made available [4].

Likert Survey. A 5-point Likert survey on LimeSurvey rated each of the 12 survey items on three constructs: *understandability*, *relevance* for agile software development, and *feasibility* in agile teams (1 = strongly disagree, 5 = strongly agree). Each cluster was introduced with a short context block paraphrasing the EU AI Act obligations it addresses, so participants rated items in context rather than in isolation.

Expert Interviews. We conducted 11 semi-structured expert interviews online via Microsoft Teams from 20 October to 25 November 2025. Participants were recruited through the professional networks of the third and fourth author of this paper and one public IT practitioner forum to diversify the sample beyond a single organization, with diversity sought along role, industry, and organization size. Sampling proceeded in two waves. The first wave covered six participants (P1–P6); we monitored category emergence after each interview and observed no new categories after the fifth interview, which we treat as theoretical saturation in the sense of Mayring [15]. We then conducted five additional interviews (P7–P11) to

■ **Table 1** Interview sample (N=11). Pseudonyms used throughout the paper. Organization sizes are banded for anonymization (10–50, 0.5k–1k, 1k–5k, 0.5k–5k, >1k, >10k, >100k employees) to preserve anonymity while retaining contextual variation. AI Act familiarity is self-reported on a five-step scale: low, low–medium, medium, medium–high, high.

ID	Role	Domain	Org. size (empl.)	AI exp. (yrs)	AI Act fam.
P1	Lead Agile Coach / RTE	Automotive	>100k	2–3	low–med
P2	Sales & Key Account Manager	IT Services / Custom Dev.	1k–5k	3–5	high
P3	Software Developer	IT / Software Development	n/a	2–4	low
P4	AI & Innovation Manager	Public Administration	>10k	1–2	high
P5	Agility Master	Railway / Critical Infrastructure	>100k	1–3	medium
P6	Former Head of IT / IT Consultant	Manufacturing / GRC	0.5k–1k	3–5	low–med
P7	Lead Agile Coach	Affiliate Marketing / SE	1k–5k	2–4	med–high
P8	Head of Transformation	Insurance	0.5k–5k	1–2	low–med
P9	CEO / Innovation Consultant	Innovation Consulting / SW Dev.	10–50	3–5	medium
P10	Engineering Manager	Affiliate Marketing / SE	1k–5k	2–3	medium
P11	Head of Digital Solutions	Industrial / E-Commerce / SAP	>1k	1–3	medium

confirm saturation and broaden contextual variation across roles, industries, and organization sizes; no new main categories emerged in this second wave. Table 1 provides an overview of the interviewee profiles and their characteristics.

The interview guide had four blocks. Blocks 1 (background and AI exposure), 2 (general assessment), and 4 (closing feedback) were identical across interviews. Block 3 was individualized per interview from the participant’s prior survey responses (the *expectation profile*) to probe divergences between low and high ratings on the same item or cluster. Interviews were held online using Microsoft Teams, were recorded with consent, transcribed automatically with Voice AI, and anonymized before analysis. The shortest interview had a duration of 18:35 minutes, whereas the longest interview lasted 43:11 minutes. The average interview duration was 30:00 minutes. This variation reflected the survey-based data, specifically the expectation profiles, on which the interviews were based. Interviews were generally shorter when there was broad overall agreement, whereas greater discrepancies required more extensive discussion. The interview guide template and the per-interview instantiations of block 3 are in the replication package.

3.3 Qualitative Data Analysis

We analyzed the transcripts with structuring qualitative content analysis following Mayring [15]. A first pass per interview produced superordinate codes; a second pass consolidated them into main and subcategories across interviews; a third pass applied the resulting coding guide (definitions, coding rules, anchor examples) to the full material. The coding guide is also available in the research protocol [4].

Coding was performed by the first author. We did not compute inter-coder reliability; the implication is discussed in Section 7. As a partial compensation, the interviewer applied in-situ communicative validation by summarizing key statements back to participants and asking for confirmation. Likert data were analyzed descriptively (means, medians, full distributions); inferential statistics are omitted because $n = 11$ does not support them.

4 The Guideline

This section presents the artifact produced by the design cycle and answers RQ1: *How can the requirements of the EU AI Act be systematically translated into actionable guidelines for the use of AI in agile software development?* We report the outcome of the traffic-light classification, the five clusters, the 12 items that constitute the guideline, and the mapping to agile elements¹. The guideline [3] is the input to the evaluation reported in Section 5.

4.1 Outcome of the Traffic-Light Classification

We assessed all 113 articles of the EU AI Act along the three dimensions (*technical, normative, and organizational*) introduced in Section 3.1. The aggregated classification resulted in 30 articles rated as high relevance, 33 as medium relevance, and 50 as low relevance. Articles classified as *high relevance* on at least one dimension entered the mapping step; articles rated *medium* or *low* on all three were retained for traceability but did not generate items. The full classification is part of the replication package (see [4]).

Three observations shape the guideline. First, high-relevance articles cluster in the high-risk obligations of Chapter III (Art. 9–15) and the transparency provisions of Chapter IV (Art. 50), with concentration on the normative and organizational dimensions. Second, the technical dimension attracts fewer high-relevance ratings than the other two; technical obligations are largely indirect, mediated by documentation and process requirements. Third, *Art. 57 regulatory sandboxes* is the only article that addresses a voluntary mechanism rather than an obligation, which justifies treating it as a separate cluster (Section 4.2).

4.2 Five Clusters Across the AI Adoption Regulation

The mapping step (Phase C, Section 3.1) linked each high-relevance article to one or more agile pain points reported by Bahi et al. [1] and Herda et al. [8]. The cluster step (Phase D) aggregated articles and pain points along thematic proximity, following the type-building qualitative content analysis according to Kuckartz and Rädiker [14], Phases 1–5. The result is five clusters that span the guideline.

Table 2 exposes the translation from regulatory anchors through mapped agile pain points to the final guideline items at cluster level. The following section provides the item-level assignments and wording.

¹ Here we reference an agile element as an agile practice or artifact in accordance with Neumann [17].

■ **Table 2** Cluster-level traceability from regulatory anchors and agile pain points to all 12 guideline items.

Regulatory anchor and obligation	Mapped agile pain point	Items
Arts. 3, 16, 26: provider and deployer roles and duties	Overlapping agile roles and unclear accountability for AI use	G1.1–G1.2
Arts. 9, 10, 14: lifecycle risk management, data governance, and human oversight	Iterative delivery can fragment risk and data-quality review and leave AI outputs without assigned human oversight	G2.1–G2.3
Arts. 11, 13, 50: documentation, instructions for use, and transparency	Minimal documentation leaves the influence of AI invisible to team members and stakeholders	G3.1–G3.2
Arts. 17, 72: quality management and post-market monitoring	Compliance monitoring and resulting process updates are not integrated into recurring team work-flows	G4.1–G4.2
Art. 57: regulatory sandboxes	Teams lack a protected setting and a traceable feedback loop for experimentation before production	G5.1–G5.3

4.3 The Twelve Items of the AI Adoption Guideline

Table 3 makes the translation chain explicit by listing, for each of the 12 items, the originating EU AI Act article or articles, the mapped agile pain point, and the resulting team-level action. Table 3 subsequently identifies the carrier artifacts through which these actions can be enacted. The pain-point formulations condense the Phase C mapping described in Section 3.1 and grounded in [1, 8]. We use the form “*teams + active verb + agile element + traceable outcome*” to keep each item actionable. Each item is short enough to fit a Definition of Done entry or a Working Agreement line.

The guideline addresses obligations on providers and deployers symmetrically. Where the EU AI Act differentiates the two roles, the items remain neutral and leave role attribution to G1.1 in Table 3.

4.4 Agile Elements Mapping

The guideline does not prescribe new agile practices. We map each item to one or more existing agile elements (Table 4) so that adoption requires adapting current artifacts rather than introducing parallel structures. This mapping is the design decision that operationalizes the agile pain point of process overhead outlined in the related work (see Section 2).

In the following, we present concrete examples of what carrier mapping might look like in practice. *Definition of Done* (G2.1, G3.1, G4.2): “If the increment uses or contains an AI component, the linked work item records data-quality evidence, the AI’s influence on the work product, and the compliance verification result; unresolved findings are escalated before release.” *Working Agreement* (G1.1, G1.2): “For each AI-related work item, the team records the responsible role, the applicable provider/deployer responsibility, and the escalation contact for legal or compliance questions.” *Sprint Review checklist* (G2.3, G3.2, G5.2): “Were relevant AI outputs reviewed by a competent human and was the result recorded? Was the AI’s influence disclosed to relevant stakeholders? If a sandbox was used, were its results and follow-up actions recorded?” These examples are illustrative carrier-level implementations, not mandatory templates or legal advice. The mapping is not bijective. Most items attach to more than one carrier, and most carriers carry more than one item. We

■ **Table 3** Items of the EU AI Act-aligned guideline for agile teams. Article numbers refer to Regulation (EU) 2024/1689.

ID	Cluster	Art.	Item
G1.1	C1 Roles & Resp.	3, 16, 26	Responsibilities for AI use are assigned explicitly and aligned with EU AI Act provider and deployer roles.
G1.2	C1 Roles & Resp.	3, 16, 26	Role assignments are documented in agile artifacts such as Working Agreements or the Definition of Done.
G2.1	C2 Risk & Quality	10	Data quality is checked and documented continuously within the team.
G2.2	C2 Risk & Quality	9	AI-related risks are identified, assessed, and documented iteratively.
G2.3	C2 Risk & Quality	14	Outputs of AI systems are reviewed regularly under human oversight.
G3.1	C3 Transparency	11, 13	Work products are documented such that the influence of AI is traceable.
G3.2	C3 Transparency	50	Transparency about AI influence is established in recurring agile events (e.g., Sprint Review).
G4.1	C4 Monitoring	17, 72	Retrospectives are used to reflect on regulatory changes and integrate them into the team's way of working.
G4.2	C4 Monitoring	17, 72	Teams establish a procedure to verify and document compliance continuously.
G5.1	C5 Sandboxes	57	AI systems are first deployed in sandbox environments before productive use.
G5.2	C5 Sandboxes	57	Sandbox results are reflected in agile events such as Sprint Reviews or retrospectives.
G5.3	C5 Sandboxes	57	Sandbox use is documented to satisfy traceability and accountability requirements.

treat this as a feature rather than a limitation: it preserves the flexibility that agile teams expect. Taken together, the traffic-light classification, the five clusters, the 12 items, and the carrier mapping form the artifact answering RQ1.

5 Findings from the Practitioner Evaluation

As stated in Section 3, we evaluated practitioners' perceptions of the guideline through a Likert-based survey and 11 semi-structured expert interviews. Section 5.1 reports the practitioner assessment along the three construct dimensions and answers RQ2. Section 5.2 reports the organizational factors that emerged from the interview analysis and answers RQ3.

5.1 Practitioner's Assessment of Understandability, Relevance, and Feasibility

The second research question, RQ2, asks: *How do practitioners assess the guidelines regarding understandability, relevance, and feasibility?* Across the 11 participants, the guideline was perceived as understandable and relevant; perceived feasibility was more dispersed and depend on the team's working context rather than on the wording of individual items. We base this assessment on the Likert ratings shown in Table 5 and on the follow-up interviews that probed deviations from the expectation profile.

■ **Table 4** Mapping of items to agile practices and artifacts.

Agile Element	Items
Working Agreement	G1.1, G1.2
Definition of Done	G1.2, G2.1, G3.1, G4.2
Backlog Refinement	G2.1, G2.2
Sprint Review	G2.3, G3.2, G5.2
Retrospective	G2.2, G4.1, G4.2, G5.2
Decision log / ADR	G3.1, G5.3
Sandbox protocol	G5.1, G5.2, G5.3

Understandability is the dimension on which the guideline is least contested. P10 framed the contrast that recurs in most interviews: condensed items are easier to act on than the full regulatory text. Only two of the 12 items attract any disagreement at all, and the dissent concentrates on terminology rather than substance, G1.2 references the *Working Agreement*, an artifact that several participants reported as unfamiliar in their own teams. P11 made the same point in a different register, asking whether the documentation duty was tied to a specific carrier or could be discharged equivalently elsewhere. The interviews position the guideline as a translation device: practitioners read it not as a new vocabulary but as an explicit version of practice that was already implicit.

Relevance disperses more than understandability and tracks the mandatory, voluntary distinction in the underlying articles. The items that operationalize Chapter III obligations of the EU AI Act, especially *human oversight* (G2.3) and *iterative risk identification* (G2.2), are rated as clearly relevant and prompted no contrary statements in the interviews. By contrast, the *continuous-compliance* item G4.1 and the *sandbox-documentation* item G5.3 attract the largest share of neutral or disagreeing responses. The qualitative material identifies two distinct mechanisms behind this. For G4.1 the criticism is one of locus: P11 rejects the retrospective as the venue for regulatory discussion, arguing that such topics “require organizational coordination rather than team-internal self-regulation”; P10 expresses the same reservation with regard to mandating regulatory reflection inside an event that is already crowded. For G5.3 the criticism is one of scope: sandbox documentation is seen as relevant where teams build AI systems but not where they merely use AI tools, a distinction (*developing AI* versus *developing with AI*) that P10 and P11 both raise independently.

Feasibility is the dimension that exposes the guideline most sharply to the working context. The single strongest objection in the entire evaluation, the only *strongly disagree* rating, is on G3.1 (documenting the influence of AI on work products), placed by P2 and accompanied by three further low ratings. The reservation is not about the principle, as transparency regarding AI influence is endorsed across the sample, but about the practical effort required in teams that have no established mechanism for documenting it. P3 captured the boundary condition that runs through the feasibility judgments: the items are realistic only when they “become part of daily work and are not perceived as an additional burden.” Where existing agile carriers (Definition of Done, Sprint Review, Retrospective) are active, integration is described as incremental; where they are weak or absent, even the most basic documentation item becomes contentious.

We summarize the distributional view and report medians and interquartile ranges in Table 5. A consistent pattern runs through the three dimensions. The guideline holds up where it formalizes practices that agile teams recognize (risk identification, human oversight,

■ **Table 5** Likert ratings per item ($N = 11$). *Mdn* = median, *IQR* = interquartile range. Five-point scale (1 = strongly disagree, 5 = strongly agree).

ID	Understandability		Relevance		Feasibility	
	<i>Mdn</i>	<i>IQR</i>	<i>Mdn</i>	<i>IQR</i>	<i>Mdn</i>	<i>IQR</i>
G1.1	4	1.0	4	1.0	4	0.0
G1.2	4	1.5	4	1.0	4	0.5
G2.1	4	1.0	4	0.5	4	0.0
G2.2	5	0.5	4	1.0	4	1.0
G2.3	5	0.0	5	1.0	5	1.0
G3.1	4	1.0	4	1.0	3	1.0
G3.2	5	1.0	4	1.0	4	1.0
G4.1	4	1.0	3	1.5	3	1.0
G4.2	5	1.0	4	1.0	4	1.0
G5.1	4	1.5	4	1.5	4	1.5
G5.2	4	1.0	4	1.0	4	0.0
G5.3	4	1.0	3	0.5	4	0.0

peer review) and meets resistance where it adds documentation or process duties that fall on the team itself (G3.1, G4.1) or addresses voluntary mechanisms (G5.3). The median rating is 4 for almost all item–dimension combinations, indicating broad agreement. The main empirical signal therefore lies in the variation of responses, which shows where practitioners’ assessments diverged.

In summary, practitioners assess the guideline as understandable and broadly relevant, with feasibility contingent on the team’s agile carriers.

5.2 Practitioner-Perceived Conditions for Adoption

Here, we answer RQ3: *Which organizational factors do practitioners perceive as influencing the adoption of our EU AI Act guideline in agile teams?* Based on the interview accounts, we identified six categories describing conditions that participants expected to support or hinder adoption. The frequencies in brackets indicate how often each category occurred in the interviews.

Function and benefit of the guideline (10/11). The guideline is positioned as an instrument of orientation rather than as a checklist. P1 described it as “a good basis to engage with AI more systematically and consciously”; P9 framed the underlying regulation as “an opportunity for quality and clarity, not as a brake.” The shared interpretation from the participants is that the guideline does not require entirely new ways of working but makes existing agile practices more explicit and links them to EU AI Act compliance. Adoption is reported to be easier where agile teams already accept this framing and harder where the guideline is read as an additional rulebook.

Implementation and integration (8/11). Whether the guideline takes hold depends on whether it is embedded into the artifacts that already structure the team’s work. For example, *Working Agreement*, *Definition of Done*, *Sprint Review*, and *Retrospective*. P3 articulated the operational boundary condition explicitly: the items work only when they “become part of daily work and are not perceived as an additional burden.” P10 reported that quality assurance for AI-generated code already runs through existing reviews, and that the value

of the guideline lies in making this explicit rather than in adding ceremonies. P8 named the converse failure mode: mere availability is not enough, and introduction needs change management and active support. Tooling and training were named as enablers across the sample; absent or implicit ownership of compliance was the most common blocker.

Roles, responsibility, and communication (8/11). The guideline is read as a redistribution of accountability, not as the assignment of compliance to a single role. Participants rejected the model of delegating AI compliance to the Product Owner, the Scrum Master, or a dedicated compliance officer in isolation. P4 emphasized that the guideline helps “make responsibilities visible, especially at the leadership level”; P9 added that formal role documentation must not replace individual ownership: “regulatory accountability cannot substitute for personal responsibility.” Two participants proposed concrete role designs: P10 suggested AI champions to coordinate compliance topics; P11 framed a defined AI responsibility as “a major opportunity for the company.” Across the sample, communication between development teams and compliance functions was named as the mechanism that converts shared responsibility into observable practice.

Agility versus regulation (7/11). Participants did not treat the guideline as a threat to agile values and principles, but they articulated the latent risk. P2 warned that strict rules “slow down the agile process”; P11 added that documentation duties “can slow down development processes.” The risk is reduced when the items extend existing events and amplified when compliance is migrated into parallel meetings. Participants endorsed integration into the Sprint Review and the Retrospective as a way to keep compliance recurrent without breaking cadence, with one notable dissent: P11 explicitly rejected the Retrospective as the venue for regulatory discussion, arguing that such topics belong to organizational coordination, not team self-regulation. This dissent maps onto the lowest relevance rating in the guideline (G4.1) and is the clearest qualitative explanation for that quantitative outlier.

Context and organizational maturity (7/11). A one-size-fits-all interpretation did not hold in practice. The same item was perceived as straightforward in one organization and difficult to implement in another, depending on organizational maturity, resource availability, and the strength of existing agile artifacts. P1 contrasted large organizations, where “the items work well,” with smaller teams that “need more individual approaches”; P9 reported the opposite from a different perspective, arguing that smaller teams may adopt the items more bindingly because they are closer to the actual work. P8 noted that the guideline “does not directly fit a non-agile context” and therefore requires translation. P10 and P11 independently raised a distinction that the guideline itself does not explicitly make: *developing AI* and *developing with AI* are different cases, and some clusters, notably C5 on sandboxes, are more relevant to the former than to the latter. According to the participants, industry was less important than organizational maturity: heavily regulated industries did not automatically find the guideline easier to adopt; what mattered was whether the regulatory function and the development teams already had an established mode of collaboration.

Participation and continuous evolution (5/11). The guideline is read as a living instrument, not as a static document. P1 summarized that competencies “cannot be conveyed once but must be built up continuously”; P4 added that practitioners themselves should be involved in revising the guideline. P7 and P10 emphasized reflection in agile events as the operational mechanism; P8 named explicit change management, training sessions, and consultation hours. Static publication of the guideline, without ownership for its ongoing revision, was identified as a likely failure mode.

Cross-category patterns

Three patterns cut across the six categories. First, organizational maturity, operationalized as the presence and active use of agile artifacts, was the most frequently reported condition associated with perceived feasibility. Agile teams with active artifacts describe integration as incremental; agile teams without them describe it as a precondition that has to be established first. Second, industry regulation builds awareness but not capacity: participants in regulated industries did not report greater ease in adopting the AI Adoption Guideline. Third, organization size shapes which carriers participants regard as realistic: smaller organizations rely on the Working Agreement as a single agile carrier, larger organizations distribute the items across more specialized artifacts. With $N = 11$ these patterns are working hypotheses for further studies, not statistical claims. Taken together, participants associated perceived adoption feasibility with three factors: collective ownership of compliance rather than delegation to a single role, embedding into existing agile events, and the maturity of the agile team's artifacts.

6 Discussion & Implications for Practice

Key Findings

Participants expected the guidelines to bridge regulation and agile delivery when teams embed them in existing artifacts and events rather than running a parallel governance track. Perceived feasibility was associated with organizational maturity and with responsibility for AI compliance being shared across the team rather than delegated to a single role. A fixed set of guidelines does not scale to organizational diversity; practitioners need a stable backbone together with explicit guidance for context-specific adaptation.

6.1 Interpretation of Findings

We interpret the interview results along three patterns that emerged consistently across the sample.

First, practitioners read the guidelines as a translation layer rather than as new policy. The function described most often across the sample is making existing practice explicit and reducing uncertainty about what Articles 9, 11, 14, and 17 of the EU AI Act [6] require inside an iteration and the applied agile practices in it. This reading is consistent with Article 96 EU AI Act [6], which positions guidelines as instruments of uniform application rather than as substantive obligations.

Second, participants associated adoption feasibility with contextual fit. The implementation and integration category was raised by 8 of 11 participants and was the most frequently reported correlate of perceived feasibility. Agile teams with established agile artifacts described integration as incremental, whereas teams without such artifacts need to create or formalize suitable structures before the AI Adoption Guideline can be embedded in day-to-day work. Industry regulation built awareness but not capacity: participants in regulated industries did not report easier adoption. P10 and P11 surfaced a distinction the guidelines themselves do not make explicit, between *developing AI* and *developing with AI*. Cluster C5 (Regulatory Sandboxes) is more pertinent to the former, while C3 (Transparency) and C4 (Monitoring) carry weight for both.

Third, responsibility for AI compliance is collective in practice but unassigned by default (8 of 11). Participants rejected delegation of EU AI Act obligations to the Product Owner, the Scrum Master, or a dedicated compliance officer in isolation. P4 stressed that the

guidelines help make responsibilities visible at the leadership level; P9 added that formal role documentation cannot substitute for individual ownership. Two participants proposed concrete designs: P10 suggested AI champions to coordinate compliance topics, P11 framed defined AI responsibility as “a major opportunity for the company.” The absence of a default owner is a known agile pain point [1]; cluster C1 (Roles and Responsibilities) encodes collective ownership in role extensions and in agile artifacts rather than in a new compliance role.

The pattern on agility versus regulation refines this picture. The risk of compliance dragging on delivery decreases when the guidelines extend existing events and increases when compliance moves into parallel meetings. One dissent on this point is informative rather than incidental: P11 explicitly rejected the Retrospective as the venue for regulatory discussion, arguing that such topics belong to organizational coordination, not team self-regulation. This single position is the clearest qualitative explanation for the lowest relevance rating in the Likert results (G4.1) and marks the limit of automatic integration into agile events.

6.2 Positioning our Contribution to the Related Work

Our results complement existing work along four areas. Todorov [25] maps AI techniques onto software engineering tasks but does not address regulatory obligations; we add the compliance layer. Bahi et al. [1] identify agile pain points around AI adoption without linking them to legal requirements; we close that link through the traffic-light mapping between high-relevance EU AI Act articles and named pain points. Kelly et al. [13] translate EU AI Act requirements into technical specifications for safety-critical AI; we cover the organizational dimension their technical framing does not reach. Herda et al. [8] report practitioner concerns from the XP 2025 workshop but stop at the level of observations; we convert comparable concerns into a practitioner-assessed set of guidelines.

Two contributions are very close to our work. Hüger et al. [10] introduce a Level of Done layer for ML projects at Cariad as the most direct prior attempt to encode EU AI Act obligations in an agile artifact; their approach targets automotive ML and is not validated with practitioners outside the case organization. Neumann et al. [18] document how teams in three German organizations improvise around the absence of clear rules and bypass emerging policies to maintain agile velocity. The present paper responds to that finding: the guidelines provide the team-level instrument those teams were missing.

The convergent point across this body of work is that EU AI Act compliance in agile contexts is an organizational problem, not a technical one. None of the prior contributions provides a practitioner-assessed set of guidelines that practitioners can adapt to their Working Agreements and Definition of Done. The artifact contribution of our paper addresses this gap. The procedure may also be adaptable to similarly structured regulations, but we did not apply it to DORA, NIS2, or the GDPR; transferability therefore remains an untested proposition for future work.

6.3 Implications for Practice

We derive seven implications. Each is grounded in interview evidence and tied to a specific thematic cluster.

L1: Extend existing agile practices before adding new ones. The strongest signal across the sample (8 of 11) is resistance to additional meetings or roles. Sprint Reviews, Retrospectives, and Backlog Refinements address the EU AI Act obligations of Articles 9 and 17 when scoped

to include AI risk and compliance reflection. P3 stated the operational condition plainly: the guidelines work only when they “become part of daily work and are not perceived as an additional burden.” Compliance migrated into parallel ceremonies was reported as the first ceremony to be dropped under delivery pressure.

L2: Anchor responsibilities in artifacts, not in job titles. Participants perceived Cluster C1 (Roles and Responsibilities) as more feasible when accountability is encoded in Working Agreements and the Definition of Done rather than assigned to a named role. P9 phrased it plainly: regulatory accountability cannot substitute for personal responsibility. Artifact-level anchoring protects the team against turnover and against the false reassurance of appointing a compliance officer with no execution mandate inside the Sprint.

L3: Calibrate guideline depth to organizational maturity. The same guideline produces different outcomes in agile teams with and without established agile practices (7 of 11). Organizations at lower maturity should adopt clusters C1 (Roles and Responsibilities) and C2 (Risk and Quality) first and defer C4 (Monitoring) and C5 (Regulatory Sandboxes) until the basic governance loop is reliable. Adopting the full set at once was reported as overwhelming. P8 named the converse failure mode for any maturity level: availability is not enough, and introduction requires change management and active support.

L4: Make AI influence visible at the artifact level. Cluster C3 (Transparency) can be implemented with low overhead through short provenance markers on user stories, pull requests, and review notes. P10 reported that quality assurance for AI-generated code already runs through existing reviews and that the value of the guideline lies in making this explicit rather than in adding ceremonies. Heavyweight documentation templates were rated as the least feasible part of the guidelines.

L5: Separate team-level execution from organizational accountability. The EU AI Act assigns obligations to providers and deployers, not to agile teams as legal entities. Teams can implement operational controls within delivery work, including risk and output reviews, documentation, traceability, monitoring, and escalation triggers. Management and organizational legal or compliance functions remain accountable for role determination, legal interpretation, risk appetite, governance structures, resources, and formal conformity decisions. The guideline therefore supports but cannot replace organizational governance or legal assessment. Before adoption, its wording and responsibilities should be reviewed with legal or compliance experts and adapted to organization-specific obligations and policies.

L6: Treat the guidelines as a living artifact. Participants read the guidelines as evolving rather than static (5 of 11). P1 noted that competencies cannot be conveyed once but have to be built up continuously. Agile teams should plan a review cadence for the guideline (e.g., on a quarterly basis) in which the guidelines are revised against retrospective input and against changes to the EU AI Act implementation acts. A frozen set loses legitimacy.

L7: Invest in AI literacy before invoking human oversight. Article 14 EU AI Act requires human oversight, but oversight without competence is ceremonial. Participants linked feasible oversight to training on model limitations, prompt design, and output assessment. For LLM-based systems, output reliability includes hallucinations, while other AI systems expose different accuracy and robustness risks. The artifact treats output reliability as a process risk through iterative risk review (G2.2), human review (G2.3), traceability (G3.1),

and pre-production sandboxing (G5.1). These controls can detect and document unreliable outputs, but cannot eliminate hallucinations or establish technical accuracy; model evaluation and technical safeguards remain necessary.

7 Threats to Validity

As with every study also ours comes with limitations. We applied the threats to validity concept [33] to mitigate the limitations and identify specific measures to optimize the quality. We discuss these threats below.

Construct Validity

The three assessment dimensions (*technical, normative, organizational*) and the three Likert constructs (*understandability, relevance, feasibility*) were defined by the first author from methodological literature and from the EU AI Act text, without external concept validation. Each Likert construct is measured by a single item per guideline, which limits construct coverage; we mitigate this by triangulating the ratings with semi-structured interviews that probe the same constructs in open form. The term *guideline* in this paper refers to a team-level translation artifact and is distinct from the Commission-issued guidelines foreseen in Art. 96 EU AI Act [6].

Because the guideline items translate selected EU AI Act obligations into concise team-level formulations, this translation may introduce interpretive bias, for example if legally relevant terms are simplified, omitted, or interpreted differently across organizational contexts. The study did not include a formal legal validation of the items by legal experts or organizational legal departments. Accordingly, the AI Adoption Guideline should be understood as an operational support instrument for agile teams, not as legal advice or as a substitute for organization-specific legal assessment.

Internal Validity

We addressed this through a documented coding guide with definitions, coding rules, and anchor examples and through a three-pass procedure (Section 3.3); both are part of the replication package. Communicative validation was partial: we used probing and summarizing during the interviews but did not return the final category system to participants for member checking. Block 3 of the interview guide was individualized from each participant's prior survey responses to target divergence between survey and interview, which may bias probing toward expected explanations; Blocks 2 and 4 remained fixed across all interviews to preserve a common baseline.

External Validity

The sample of 11 participants may not be representative. Recruitment proceeded through the professional networks of co-authors, one public IT practitioner forum, and targeted primarily agile and managerial roles. The sample includes one software developer (P3), but no data scientists, legal experts, or compliance specialists; technical and legal/compliance perspectives are therefore underrepresented. The EU AI Act is in its early implementation phase, with obligations for high-risk systems applying from 2 August 2026, and the regulatory context is European; transfer to other jurisdictions or to later enforcement phases is not warranted by our data. Cluster C5 addresses regulatory sandboxes under Art. 57, which are voluntary; feasibility ratings for C5 generalize only to organizations that actively consider sandbox participation.

Conclusion Validity

The Likert data are ordinal, and the sample size precludes inferential statistics; we report medians and interquartile ranges and do not test differences across guidelines or constructs. The qualitative analysis carries the interpretation margin inherent to structuring content analysis; the cross-case patterns in Section 5.2 are working hypotheses, not statistical claims. This study assesses perceived understandability, relevance, and feasibility; it does not observe guideline adoption in the field or test whether the artifact improves compliance readiness. We draw no causal conclusions about adoption.

Reliability

The replication package [4] covers the traffic-light mapping, the Article-to-cluster-to-guideline mapping, anonymized Likert raw data, the expectation profile, the interview guide, the coding guide, and the anonymized analysis matrix.

8 Conclusion & Future Work

This paper addresses how agile teams can discharge EU AI Act obligations without disrupting iterative delivery practices. Three findings carry our contribution. First, the regulation's obligations can be translated into actionable guidelines for agile teams through a reproducible procedure that combines a three-dimensional article assessment, a traffic-light classification, and a mapping to reported agile pain points. Second, practitioners assessed the resulting 12 guidelines items in five clusters as understandable and relevant; perceived feasibility of adoption varied with organizational maturity and the strength of existing agile artifacts. Third, participants associated perceived adoption feasibility with three organizational factors: collective ownership of compliance, integration into existing agile events, and treatment of the guideline as a living artifact rather than a static document.

Three directions extend this work. A larger and more diverse practitioner sample is needed to test whether the cross-category patterns observed generalize across industries, organization sizes, and maturity levels. Longitudinal pilot studies inside the industry partner organization would shift the evidence base from perceived to actual adoption, traceable in Working Agreements, the Definition of Done, and the carrier artifacts of clusters C1–C5. The translation procedure uses analytical steps that are not specific to the EU AI Act, but its transferability has not yet been empirically evaluated. Applying it to DORA, NIS2, and the GDPR would test which elements are transferable and which require regulation-specific adaptation.

Data Availability Statement

The replication package for this study is publicly available [4]. It includes the traffic-light classification, the article-to-cluster-to-guideline mapping, the expectation profiles, the customized interview guide based on the expectation profiles, the coding guide & book, and the anonymized analysis matrix. The survey raw data and interview transcripts are not publicly available in order to protect participant confidentiality and to comply with data protection requirements. Access to additional anonymized materials may be provided by the corresponding author upon reasonable request, subject to ethical and legal constraints.

We also made the guideline available as a companion document [3].

References

- 1 Asmae Bahi, Jihane Gharib, and Youssef Gahi. Integrating generative AI for advancing agile software development and mitigating project management challenges. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 15(3):54–61, 2024. doi: 10.14569/ijacsa.2024.0150306.
- 2 Jürgen Bortz and Nicola Döring. *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. Springer Medizin, Heidelberg, 4. vollständig überarbeitete edition, 2006.
- 3 Schrader Dennis, Eva-Maria Schön, Henning Fritzemeier, and Michael Neumann. Eu ai act-aligned guidelines for agile software development, 2026. Available at: <https://doi.org/10.5281/zenodo.21804652>.
- 4 Schrader Dennis, Eva-Maria Schön, Henning Fritzemeier, and Michael Neumann. Research protocol, 2026. Available at: <https://doi.org/10.5281/zenodo.21804701>.
- 5 Emelie Engström, Margaret-Anne Storey, Per Runeson, Martin Höst, and Maria Teresa Baldassarre. How software engineering research aligns with design science: a review. *Empirical Software Engineering*, 25(4):2630–2660, 2020. doi:10.1007/s10664-020-09818-7.
- 6 European Union. Regulation (EU) 2024/1689 of the european parliament and of the council of 13 june 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union, L series, 2024. 12 July 2024. URL: <https://eur-lex.europa.eu/eli/reg/2024/1689/oj>.
- 7 Sergio Genovesi. Introducing an AI governance framework in financial organizations: Best practices in implementing the EU AI Act (practitioner track). In *Symposium on Scaling AI Assessments (SAIA 2024)*, volume 126 of *Open Access Series in Informatics (OASICS)*, pages 9:1–9:7. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2025. doi:10.4230/OASICS.SAIA.2024.9.
- 8 Tomas Herda, Victoria Pichler, Zheyang Zhang, Pekka Abrahamsson, and Geir K. Hanssen. Ai and agile software development: From frustration to success xp2025 workshop summary. In *Proceedings of the 26th International Conference on Agile Software Development – Workshops*, pages 3–13, Cham, 2025. Springer Nature Switzerland. doi:10.1007/978-3-032-05799-0_1.
- 9 Alan R. Hevner, Salvatore T. March, Jinsoo Park, and Sudha Ram. Design science in information systems research¹. *Management Information Systems Quarterly*, 28(1):75–106, March 2004. doi:10.2307/25148625.
- 10 Fabian Hüger, Alexander Poth, Andreas Wittmann, and Roland Walgenbach. An approach to the instantiation of the EU AI Act: A Level of Done derivation and a case study from the automotive domain. In *Systems, Software and Services Process Improvement – 30th European Conference, EuroSPI 2023, Grenoble, France, August 30 – September 1, 2023, Proceedings*, volume 1890 of *Communications in Computer and Information Science*, pages 111–123. Springer, 2023. doi:10.1007/978-3-031-42307-9_9.
- 11 Isabelle Hupont, David Fernández-Llorca, Sandra Baldassarri, and Emilia Gómez. Use case cards: A use case reporting framework inspired by the European AI Act. *Ethics and Information Technology*, 26(2):19, 2024. doi:10.1007/s10676-024-09757-7.
- 12 International Organization for Standardization. ISO/IEC 42001:2023 – Information technology – Artificial intelligence – Management system. <https://www.iso.org/standard/81230.html>, 2023. ISO/IEC 42001:2023.
- 13 Jessica Kelly, Shanza Ali Zafar, Lena Heidemann, João-Vitor Zacchi, Delfina Espinoza, and Núria Mata. Navigating the EU AI Act: A methodological approach to compliance for safety-critical products. In *2024 IEEE Conference on Artificial Intelligence (CAI)*. IEEE, 2024. doi:10.1109/CAI59869.2024.00159.
- 14 Udo Kuckartz and Stefan Rädiker. *Qualitative Inhaltsanalyse – Methoden, Praxis, Umsetzung mit Software und künstlicher Intelligenz*. Beltz Juventa, 6. edition, 2024.
- 15 Philipp Mayring. *Qualitative Content Analysis: A Step-by-Step Guide*. SAGE Publications Ltd, London, 1. edition, 2022.

- 16 Jakob Mökander, Maria Axente, Federico Casolari, and Luciano Floridi. Conformity assessments and post-market monitoring: A guide to the role of auditing in the proposed European AI regulation. *Minds and Machines*, 32(2):241–268, 2022. doi:10.1007/s11023-021-09577-4.
- 17 Michael Neumann. Towards a taxonomy of agile methods: The tree of agile elements. In *2021 9th International Conference in Software Engineering Research and Innovation (CONISOFT)*, pages 79–87, 2021. doi:10.1109/CONISOFT52520.2021.00022.
- 18 Michael Neumann, Lasse Bischof, Nic Elias Hinz, Abdullah Altun, Luca Stockmann, Dennis Schrader, Ana Carolina Ahaus, Erim Can Demirci, Benjamin Gabel, Maria Rauschenberger, Philipp Diebold, Henning Fritzscheier, and Adam Przybytek. Between policy and practice: GenAI adoption in agile software development teams. In *Proceedings of the 27th International Conference on Agile Software Development*, volume 578 of *Lecture Notes in Business Information Processing*, pages 290–308, Cham, 2026. Springer. doi:10.1007/978-3-032-22375-3_18.
- 19 Mirko Perkusich, Danylo Albuquerque, Allysson Alex Araújo, Matheus Paixão, Rohit Gheyi, Marcos Kalinowski, and Angelo Perkusich. Adoption of large language models in Scrum management: Insights from brazilian practitioners. In *Agile Processes in Software Engineering and Extreme Programming – 27th International Conference, XP 2026, São Paulo, Brazil, April 13–17, 2026, Proceedings*, volume 578 of *Lecture Notes in Business Information Processing*, pages 255–273, Cham, 2026. Springer. doi:10.1007/978-3-032-22375-3_16.
- 20 Daniel Planötscher, Xiaofeng Wang, Mahum Adil, Peggy Gregory, and Sepideh Larimian. Human-AI collaboration in software development activities: Perspectives of agile practitioners. In *Proceedings of the 27th International Conference on Agile Software Development*, volume 578 of *Lecture Notes in Business Information Processing*, pages 274–289, Cham, 2026. Springer. doi:10.1007/978-3-032-22375-3_17.
- 21 Daniel Russo, Sebastian Baltes, Niels van Berkel, Paris Avgeriou, Fabio Calefato, Beatriz Cabrero-Daniel, Gemma Catolino, Jürgen Cito, Neil Ernst, Thomas Fritz, Hideaki Hata, Reid Holmes, Maliheh Izadi, Foutse Khomh, Mikkel Baun Kjærgaard, Grischa Liebel, Alberto Lluch Lafuente, Stefano Lambiase, Walid Maalej, Gail Murphy, Nils Brede Moe, Gabrielle O’Brien, Elda Paja, Mauro Pezzè, John Stouby Persson, Rafael Prikladnicki, Paul Ralph, Martin Robillard, Thiago Rocha Silva, Klaas-Jan Stol, Margaret-Anne Storey, Viktoria Stray, Paolo Tell, Christoph Treude, and Bogdan Vasilescu. Generative AI in software engineering must be human-centered: The Copenhagen manifesto. *Journal of Systems and Software*, 216:112115, 2024. doi:10.1016/j.jss.2024.112115.
- 22 Chalisa Veasommai Sillberg, José Siqueira De Cerqueira, Pekka Sillberg, Kai-Kristian Kemell, and Pekka Abrahamsson. The EU AI Act is a good start but falls short. In *Proceedings of the 15th International Conference on Software Business*, volume 539 of *Lecture Notes in Business Information Processing*, pages 114–130. Springer, 2025. doi:10.1007/978-3-031-85849-9_10.
- 23 Francesco Sovrano, Emmie Hine, Stefano Anzolut, and Alberto Bacchelli. Simplifying software compliance: AI technologies in drafting technical documentation for the AI Act. *Empirical Software Engineering*, 30:91, 2025. doi:10.1007/s10664-025-10645-x.
- 24 Elham Tabassi. Artificial Intelligence Risk Management Framework (AI RMF 1.0). Technical Report NIST AI 100-1, National Institute of Standards and Technology, 2023. doi:10.6028/NIST.AI.100-1.
- 25 Plamen G. Todorov. The application of artificial intelligence in software engineering. *International Journal of Advanced Multidisciplinary Research and Studies*, 2(5):835–842, 2022.
- 26 Michael Tuape, Yohannes Garbielmichael, and Jussi Kasurinen. Architecting trust: Designing human oversight and accountability for AI-driven software engineering under the EU AI Act. In *Proceedings of the 13th International Conference in Software Engineering Research and Innovation (CONISOFT 2025)*. IEEE, 2025. doi:10.1109/CONISOFT66928.2025.00048.
- 27 Rasmus Ulfesnes, Nils Brede Moe, Viktoria Stray, and Marit Skarpen. Transforming software development with generative AI: Empirical insights on collaboration and workflow. In Anh Nguyen-Duc, Pekka Abrahamsson, and Foutse Khomh, editors, *Generative AI for Effective Software Development*, pages 219–235. Springer, 2024.

- 28 Matthias Wagner, Markus Borg, and Per Runeson. Navigating the upcoming European Union AI Act. *IEEE Software*, 41(1):19–24, 2024. doi:10.1109/MS.2023.3322913.
- 29 Matthias Wagner, Rushali Gupta, Markus Borg, Emelie Engström, and Michal Lysek. AI Act high-risk requirements readiness: Industrial perspectives and case company insights. In *Proceedings of the 25th International Conference on Product-Focused Software Process Improvement*, Lecture Notes in Computer Science. Springer, 2024. doi:10.1007/978-3-031-78392-0_5.
- 30 Matthias Wagner, Marie-Therese Wittmann, and Markus Borg. A critical look at the EU AI Act’s requirements and affected systems: Who must explain what? In *2025 IEEE 33rd International Requirements Engineering Conference Workshops (REW)*, pages 494–503, Valencia, Spain, 2025. IEEE. doi:10.1109/REW66121.2025.00074.
- 31 Salla Westerstrand. Fairness in AI systems development: Beyond EU AI Act compliance. In *Proceedings of the 15th International Conference on Software Business*, volume 539 of *Lecture Notes in Business Information Processing*, pages 99–113. Springer, 2025. doi:10.1007/978-3-031-85849-9_9.
- 32 Roel J. Wieringa. *What Is Design Science?*, pages 3–11. Springer Berlin Heidelberg, Berlin, Heidelberg, 2014. doi:10.1007/978-3-662-43839-8_1.
- 33 Claes Wohlin, Per Runeson, Martin Höst, Magnus C. Ohlsson, Björn Regnell, and Anders Wesslén. *Experimentation in Software Engineering*. Springer, Berlin, Heidelberg, 2. edition, 2012. doi:10.1007/978-3-642-29044-2.