RESEARCH IN PROGRESS

Evaluating an AI Act Regulatory Sandbox Framework Against Pilot Cases: Towards a Methodologically Grounded Three-Round Process Framework

Jukka Heikkilä,[1] Altti Lagstedt,[2] Seppo Heikura,[1] Anastasiia Dardykina[3]

[1] University of Turku, Turku School of Economics, Turku, Finland
jups@utu.fi, seppo.heikura@utu.fi
[2] Haaga-Helia University of Applied Science, Helsinki, Finland
altti.Lagstedt@haaga-helia.fi
[3] University of Vaasa, School of Accounting and Law, Vaasa, Finland
anastasiia.dardykina@uwasa.fi

This research-in-progress paper advances out tentative AI Regulatory Sandbox Framework (AIRSF) in two directions. Empirically, it evaluates 46 pilot cases and with five workshops of Nordic national/regional AI ecosystems. Conceptually, it argues that the AI Act's dual objective of innovation and compliance can only be operationalised in a business-friendly manner if grounded in participatory, situation-adaptive method development with dispute resolution procedures on controversial issues. Earlier we have identified four methodological steps: stakeholder mapping, conflict surfacing, iterative KPI-building, institutionalised conflict resolution, and show their concrete design consequences for a three-round sandbox related sub-processes: (1) a revised use case and risk category definition procedures with company-paced entry; (2) multi-authority evaluation with agile guidance and co-ordination; (3) post-market surveillance options with iterative re-entry. Preliminary findings surface three structural challenges: entry friction for determining use case purpose, borderline high-risk cases, capabilities altering the AI System design and purpose while in sandbox, and a post-market visibility gap in Round 3.

DOI
https://doi.org/10.18690/um.fov.4.2026.71

ISBN
978-961-299-152-4

Keywords:
AI act, regulatory sandbox, stage-gate model, participatory method engineering, triple helix, post-market surveillance, SMEs

University of Maribor Press

1 Introduction and method

The EU AI Act (AIA) mandates at least one AI Regulatory Sandbox (AIRSF) in each Member State by August 2026, with a dual objective encoded in Art. 57(9): foster innovation and ensure regulatory compliance (EC, 2025a). For AI providers and deployers, particularly SMEs, the sandbox is potentially the difference between market access and exclusion. The AIRSF situates itself within the IS research stream on regulatory governance infrastructures (Tilson et al., 2010) addressing the identified gap that existing IT governance frameworks address internal organisational alignment but rarely incorporate dynamic, multi-party conflict resolution (Sambamurthy & Zmud, 1999). Art. 57(9)'s objectives (b) and (d) structurally serve regulator interests while (a) and (c) serve company interests (Dardykina et al., 2026), an explicit conflict-resolution method is a design necessity.

Building on the AIRSF presented at Bled 2025 (Heikkilä et al., 2025), a seven-gate Stage-Gate model as shared boundary object for sandbox co-design, this paper evaluates the tentative framework against 46 pilot cases and five workshops. The AIA's dual objective is only achievable if sandbox design is grounded in participatory, situation-adaptive method engineering. Without explicit conflict resolution, implementations default to regulator-centric or company-centric designs, neither of which realises the Triple Helix ideal (Etzkowitz & Leydesdorff, 2000) the AIA requires. We address three research questions: RQ1 How do the AIRSF's design requirements hold up against pilot case evidence? RQ2 What methodological grounding is necessary to resolve Art. 57(9)'s conflicting interests? RQ3 What are the eligibility/entry criteria for AI Systems in the pre-entry phase and for substantially changed systems post-market? The pilot case set comprises 46 AI providers, deployers, and National Competent Authorities (NCAs) across the EU, selected through selective sampling prioritising geographic spread, actor-role diversity, and system-type coverage, supplemented by 4 regional workshops combining Design Science Research with Action Research (Iivari & Venable, 2009).

2 From Triple Helix to Revising Tentative model, preliminary findings

Dardykina et al. (2026) identify four methodological steps for sandbox governance: (1) map all affected parties and interests; (2) surface and negotiate conflicts via structured scenario methods (Godet, 2000); (3) build solutions iteratively with

transparent KPIs; and (4) establish institutionalised conflict-resolution mechanisms. These steps correspond to three theoretical groundings. Participatory method engineering (Checkland, 2000) provides the conflict-surfacing tools for Steps 2 and 4. Design Science Research (Van Aken & Romme, 2009; Baskerville et al., 2018; Hevner et al., 2004) supplies the iterative evaluation logic for Steps 1 and 3, enabling agile refinement of compliance criteria without full process restart. The Triple Helix model (Etzkowitz & Leydesdorff, 2000) provides the institutional coordination frame, repositioning the sandbox from a one-directional checkpoint into a middle-out learning space where business, authority, and innovation ecosystem actors co-produce compliance. Without this integration, a purely process-oriented sandbox remains a sequential inspection station rather than a generative institution for trustworthy AI governance.

Our suggested approach serves a triple helix activity among the businesses, value adding parties of the innovation ecosystem and the authorities during and after the sandbox. We use the following simplified figure (Figure 1) of the tentative process (Heikkilä et al., 2025) to condense the stage-gate model to the minimum (c.f. Heikkilä, 2025; Heikkilä & Lagstedt, 2025).

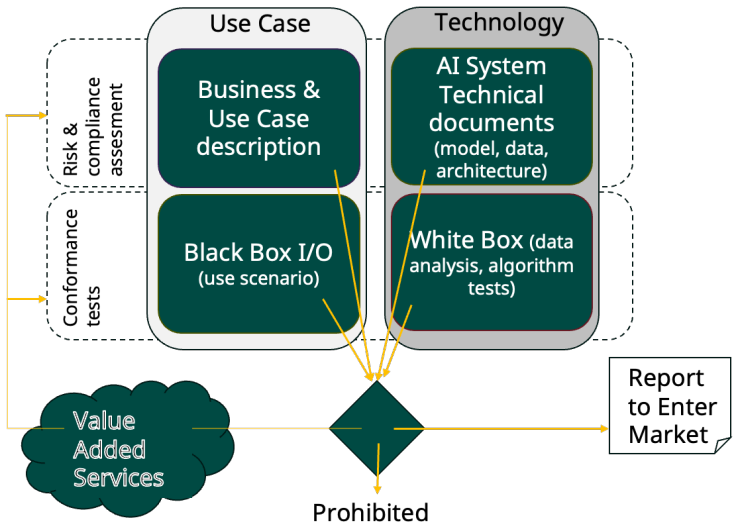

Figure 1: The Simplified Sandbox & Value Added Services Representation (Use Case × Technology evaluation matrix)

Source: Authors

3 Round 1: Use Case Definition and Company-Paced Entry

Building on Heikkilä et al. (2025), sandbox participation begins with AI provider self-evaluation (EC, 2025b) covering use case documentation, initial AIA compliance, and risk categorisation. Pilot cases reveal high demand for guidance on borderline-risk systems (medical devices, biometric tools, digital services), where a minor design and use-case change can shift the risk category, particularly for SMEs facing overlapping regulations (GDPR, Medical Device regulation). System maturity must exceed Technology Readiness Level 3 (TRL 3) (Martínez-Plumed et al., 2021), with a clearly defined use case, training and input data process, documented compliance attempts, and a credible go-to-market strategy, agentic AI's dynamism demands additional scrutiny (Yousefi et al., 2025). Company and business validity, adequate funding, viable business model, and value chain, are also assessed, while cross-domain AI systems (e.g. object identification across healthcare and autonomous driving) require deep use-case and technical analysis (Fig. 1) supported by innovation ecosystem services.

Pilot analysis reveals a more interactive guidance process than anticipated: company representatives are supported by legal, technical, and business experts to map systems along the AIRSF's two-dimensional use case typology (number of models × domain breadth) and produce a preliminary risk estimate (Use Case column, Fig. 1). This has prompted a key AIRSF revision: Gate 1 has been split into a lightweight 'pre-entry' sub-process with company-paced documentation support, distinct from the formal NCA eligibility decision, reducing entry friction without compromising evaluation integrity.

4 Round 2: Multi-Authority Evaluation of Compliance with/without Conformance testing

Pilot evidence shows the number of involved authorities can grow to tens (e.g. GDPR, IPR, Cyber Security, Product safety and its Quality Systems), raising an open question of whether product-safety testing with AI components should be embedded in the sandbox or treated as an auxiliary activity, with co-ordination potentially residing with innovation ecosystem parties providing services to start-ups and SMEs. Three design consequences follow. First, entry must include a rapid self-assessment followed by a formal NCA eligibility decision with transparent criteria,

in days rather than weeks. Second, testing must accommodate agile, in-turn commentary and development, since a sandbox functioning as a final-stage inspection station is incompatible with the fail-fast model of most start-ups (Dardykina et al., 2026). This may require the use of some external experts at appropriate stages of the process. Third, re-entry after modification must not require full process restart, only the changes requested by the supervisory body are re-verified. These findings have prompted revision of the AIRSF Round 2 gate structure: a rolling-wave evaluation protocol now replaces the original single compliance checkpoint, allowing in-turn authority commentary and partial re-evaluation of changed components without restarting the full multi-authority review.

5 Post-Market Surveillance and Re-Entry to the sandbox.

The third sub-process extends the AIRSF beyond the exit report. Three further sub-processes are required: (a) continuous monitoring of the deployed system against the exit-report compliance baseline, with an AI Educational Center (Dardykina et al., 2026) as the institutional repository for accumulated learning; (b) detection and flagging of changes crossing the AIA Art. 83 'substantial modification' threshold, which are in a clear case a change of the purpose of the AI System, change of data, model or both (c) conditional re-entry at the gate appropriate to the modification, depending on the exit state for revision, or value-added services expertise and guidance. The open questions concern who operationalises the trigger criteria and who bears observation responsibility – questions that methodological step 4 demands be explicitly assigned, since provider incentive structures do not reliably generate timely self-reporting of compliance-relevant changes. This gap has prompted revision of the AIRSF's Round 3 design: explicit accountability assignments for post-market observation are now embedded as a negotiated output of Round 2, with the AI Educational Center (Dardykina et al., 2026) designated as the institutional repository for monitoring and re-entry triggering.

Finally, NCAs increasingly rely on RegTech tools to guide, document, and analyse AI systems, tools that companies already use. As the number of market solutions grows, Market Surveillance Authorities (MSAs) will need adapted versions of these same tools for effective surveillance.

6 Discussion and Open Questions

Table 1 shows the round-specific open questions and their methodological stakes. Two key tensions frame the discussion. First, a specification–development mismatch: the AIA expects systems to be well-defined at sandbox entry, whereas the target actors' iterative practices assume the opposite. This creates a clash between regulatory epistemology (stable systems enable risk classification) and innovation epistemology (specification emerges through iteration). The Round 2 agile-compatibility measures respond to this but implementing them requires the participatory method's ability to surface such conflicts, not process design alone.

Secondly, the dual-objective tension cannot be solved merely by allocating functions across the three rounds unless institutional roles in each are explicitly negotiated. The paper's methodological contribution is to show that the AIA's Triple Helix ambition requires a participatory method as its operational base; without it, the sandbox fails both companies and regulators, an outcome by 46 pilot cases. The full paper will present the thematic analysis, a revised AIRSF diagram with the three-round structure and agile re-entry pathways, and design propositions refined through stakeholder co-design.

Table 1: Open design and governance questions by round

Round	Open Question1	Methodological Stake
R1	Separate deployer entry lane or provider-led process? (AIA Art. 25 accountability)	Stakeholder mapping; all affected parties identified
R1	What business significance criteria determine priority access for borderline high-risk cases? Can 3–4 day NCA assessment be operationalised?	Company-paced entry; SME problems tackled in advance
R2	Is notified-body conformance testing inside or outside the sandbox boundary? (cf. Fig. 1 conformance track)	Institutional scope; avoiding redundant evaluation silos
R2	How to accommodate agile/fail-fast cycles within the gate structure without undermining evaluation integrity?	Rolling-wave planning; sandbox as early engagement, not final checkpoint
R3	What operationalises AIA Art. 83 'substantial modification' as an observable re-entry trigger?	Legal predictability of change of purpose; re-entry without full process restart in case of tech/data change.
R3	Who observes and flags changed AI behaviour post-deployment — provider, deployer, NCA, or MSA?	Accountability assignment; institutionalised conflict resolution.

References

Baskerville, R., Baiyere, A., Gregor, S., Hevner, A., & Rossi, M. (2018). Design science research contributions: Finding a balance between artifact and theory. In *Journal of the Association for Information Systems* (Vol. 19, Number 5, pp. 358–376). Association for Information Systems. https://doi.org/10.17705/1jais.00495

Checkland, P. (2000). Soft Systems Methodology: A Thirty Year Retrospective a. *Systems Research and Behavioral Science Syst. Res*, *17*, 11–58. https://doi.org/10.1002/1099-1743(200011)17:1

Dardykina, A., Heikkilä, J., Lagstedt, A., Heikura, S., & Jokinen, J. (2026). *Who is Really Playing in the Sandbox?* https://cybersecnatlab.it/wp-

EC. (2025a). *AI act*. Https://Digital-Strategy.Ec.Europa.Eu/En/Policies/Regulatory-Framework-Ai.

EC. (2025b). *AI Act Single Information Platform*. Https://Ai-Act-Service-Desk.Ec.Europa.Eu/En.

Etzkowitz, H., & Leydesdorff, L. (2000). The dynamics of innovation: from National Systems and "Mode 2" to a Triple Helix of university-industry-government relations. In *Research Policy* (Vol. 29). www.elsevier.nlrlocatereconbase

Godet, M. (2000). *The Art of Scenarios and Strategic Planning: Tools and Pitfalls*. http://www.cnam.fr/lips/,

Heikkilä, J. (2025). *A presentation at ASCEND Confrence under the Danish Presidency of the Council of the EU: Strategy to Action: Closing the European Innovation and Technology Gap*.

Heikkilä, J., & Lagstedt, A. (2025). *A tutorial at ECAI*. https://sites.google.com/view/sandbox-ecai2025-tutorial

Heikkilä, J., Lagstedt, A., Heikura, S., Kaakkurivaara, E., Dardykina, A., & de Chardin, A. T. (2025). *From Simple Sandbox Process to Regulatory Sandbox Framework: Serving the Dual Objectives of AI Regulation*. 643–660. https://doi.org/10.18690/um.fov.4.2026.40

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). DESIGN SCIENCE IN INFORMATION SYSTEMS RESEARCH 1. In *Design Science in IS Research MIS Quarterly* (Vol. 28, Number 1).

Iivari, J., & Venable, J. R. (2009). *Association for Information Systems AIS Electronic Library (AISeL) Action research and design science research-Seemingly similar but decisively dissimilar Recommended Citation*. http://aisel.aisnet.org/ecis2009/73

Martínez-Plumed, F., Gómez, E., & Hernández-Orallo, J. (2021). Futures of artificial intelligence through technology readiness levels. *Telematics and Informatics*, *58*. https://doi.org/10.1016/j.tele.2020.101525

Sambamurthy, V., & Zmud, R. W. (1999). *ARRANGEMENTS FOR INFORMATION TECHNOLOGY GOVERNANCE: A THEORY OF MULTIPLE CONTINGENCIES^*.

Tilson, D., Lyytinen, K., & Sørensen, C. (2010). Digital infrastructures: The missing IS research agenda. *Information Systems Research*, *21*(4), 748–759. https://doi.org/10.1287/isre.1100.0318

Van Aken, J. E., & Romme, G. (2009). Reinventing the future: Adding design science to the repertoire of organization and management studies. *Organisation Management Journal*, *6*(1), 5–12. https://doi.org/10.1057/omj.2009.1

Yousefi, Y., Billi, M., & Rotolo, A. (2025). *Agentic AI: An EU AI Act Paradigm Shift?* https://doi.org/https://dx.doi.org/10.2139/ssrn.5731424