Qeios

Peer Review

Review of: "Decision Support or High-Risk? Classifying AI Workforce-Intelligence Systems Under the EU AI Act"

Rajeew Vishvakarma[1]

1. Information Technology, Infosys (India), Bengaluru, India

Overall Assessment

This manuscript explores the distinction between workforce intelligence systems, which provide aggregate and strategic labor market insights, and high-risk employment AI systems under the EU AI Act. This is important for researchers and practitioners in AI governance, HR technology, workforce planning, responsible AI, and regulatory compliance. This study advances beyond doctrinal analysis to translate regulatory concepts into system design and governance practices. The four-archetype taxonomy, five architectural properties, reference architecture, operational checklist, and RACI model provide a practitioner-oriented framework. However, I recommend some revisions. Although the central argument is promising, the legal classification logic requires greater precision. The manuscript should clearly distinguish between the factors determining regulatory classification and those improving governance, auditability, or compliance readiness.

Major Comments

1. The regulatory implementation timeline should be updated

The manuscript states that the bulk of the high-risk regime will become applicable on August 2, 2026. However, this is no longer accurate under the consolidated AI Act following the 2026 amendments.

For AI systems classified as high-risk pursuant to Article 6(2) and Annex III—including employment-related systems—the relevant Chapter III Sections 1–3 requirements will apply from **December 2, 2027**. Systems classified under Article 6(1) and Annex I follow a later date of August 2, 2028.

doi.org/10.32388/MMDWEP

Because the manuscript repeatedly presents itself as guidance for phased implementation, this date should be corrected consistently in the abstract, introduction, implications, and conclusion.

2. The classification framework should use a clearer two-stage legal test

This study places considerable weight on Article 6(3), describing it as central to distinguishing decision support from high-risk AI. However, Article 6(3) is a derogation applicable to a system that is referred to for the first time **in Annex III**. Therefore, logical order is important.

I suggest restructuring the framework around two sequential questions:

Stage 1: Annex III scope.

Is the system's intended purpose one of the employment uses identified in Annex III, point 4?

Stage 2: Article 6(3) derogation stage.

If the system falls within Annex III, does it nevertheless satisfy the conditions for being treated as non-high-risk under Article 6(3)?

Annex III point 4 specifically addresses systems intended for recruitment or selection and systems used to make or support specified decisions concerning employment relationships, task allocation, or worker monitoring and evaluation.

This distinction is especially important for Archetypes C and D. A genuinely strategic labor market intelligence platform may be **outside Annex III altogether**, rather than an Annex III system that subsequently qualifies for the Article 6(3) derogation. The current framework occasionally blurs the distinction between these two legally distinct routes.

A decision tree separating these pathways would substantially improve this manuscript.

3. "Architecture determines classification" should be qualified

The manuscript's central proposition that classification depends on architecture rather than data type is useful but somewhat overstated. Architecture can constrain how a system is used, reduce the likelihood of individual profiling, provide evidence of the intended purpose, and prevent scope creep. Therefore, it strongly contributes to defensibility.

However, the Act's employment classifications are fundamentally expressed in terms of systems **intended** for specified purposes.

I recommend reframing the central proposition as follows:

Classification principally depends on the intended purpose and deployment context, whereas architecture can constrain, evidence, and preserve the intended purpose.

Such wording would preserve the paper's engineering contribution while aligning the framework more precisely with the regulatory structure.

4. The five architectural properties do not have equal classification significance

The five proposed properties are valuable; however, the manuscript should distinguish between the following:

- **classification-relevant properties**, and
- **governance and compliance-supporting properties**.

The aggregation boundary and prevention of individual profiling are highly relevant to the classification. The AI Act expressly provides that an Annex III system performing the profiling of natural persons is always considered high-risk.

Decision proximity and the intended purpose are similarly significant.

In contrast, determinism, reproducibility, provenance logging, and version control are primarily governance, auditability, robustness, and evidentiary controls. A deterministic system that ranks job applicants could clearly remain high-risk, while a probabilistic system producing aggregate labor market forecasts might remain outside Annex III.

This distinction is important because the current wording could lead practitioners to believe that implementing all five controls meaningfully changes the legal classification of a system. I suggest revising the framework to explicitly show which properties affect the classification analysis and which properties strengthen compliance or defensibility after classification.

5. Human oversight and read-only interfaces should not be treated as strong evidence of non-high-risk status by themselves

The manuscript argues that explain-only outputs, the absence of system write-back, and human-in-the-loop review reinforce decision-support characterization. These are sensible safeguards, but the analysis should recognize that a recommendation can materially influence an employment decision even when a human formally makes the final decision.

A system need not autonomously execute employment actions to have a substantial decision-making influence. Consequently, this study should examine **decision coupling** rather than simply whether an API or technical write-back mechanism exists.

For example, a read-only system whose recommendations are routinely followed by workforce planners could potentially exert greater practical influence than a technically integrated system whose outputs are only one minor factor among many.

Therefore, I recommend adding questions such as the following:

- How frequently are the system recommendations followed?
- Are managers required to justify their departures from the model?
- Does the system determine which cases require human attention?
- Are system recommendations systematically translated into restructuring, promotion, hiring, or redundancy actions?

This would considerably strengthen sociotechnical analysis.

6. "Public" data should not be treated as equivalent to aggregate or non-personal data

The paper appropriately emphasizes the aggregation boundary, but several passages group "public and aggregate" labor market information together. These concepts should be kept distinct.

Public sources may still contain information relating to identifiable natural persons. The manuscript refers to leadership and executive movements that may involve named individuals. Therefore, such data require separate analysis from genuinely aggregated information concerning roles, skill demand, locations, or labor market populations.

The architecture should ideally distinguish the following:

public-source personal data → aggregation/transformation → non-individual analytical construct → strategic output

rather than suggesting that public-source status provides regulatory insulation.

This clarification would also strengthen the manuscript's discussion of the risks of profiling and re-identification.

7. The design-science contribution requires stronger evaluation

This manuscript explicitly presents its taxonomy, five properties, reference architecture, checklist, and practice layer as design-science artifacts. However, empirical evaluation has been largely deferred to future research.

This represents an important methodological limitation of the study. At present, statements that the five properties "materially affect classification and defensibility" are primarily analytically derived propositions rather than empirically evaluated findings.

The paper would be significantly stronger if the author performed at least one evaluation at the present stage. Possible approaches include the following:

- structured expert interviews with AI governance, HR technology, employment law, and ML specialists;
- A Delphi evaluation of the taxonomy and five properties;
- application of the framework to several realistic workforce intelligence scenarios;
- comparison of classifications reached by independent reviewers;
- a structured FEDS evaluation of its utility, completeness, and usability.

Even a modest expert validation exercise would materially increase confidence in the proposed artifacts.

8. The manuscript should engage with the Commission's 2026 draft high-risk classification guidance

The manuscript describes the Commission's guidance on distinguishing high-risk from non-high-risk systems, largely as an anticipated future development. However, the European Commission published **draft guidelines on the classification of high-risk AI systems on May 19, 2026**, including practical examples designed to assist providers and deployers in applying Article 6.

Given that the entire manuscript concerns the application of Article 6, these guidelines should be incorporated into the analysis of the study. Although they remain draft guidance, they are directly relevant to the manuscript's core research problem.

Minor Comments

First, the manuscript is generally well-written; however, some sections are repetitive. The distinction between aggregate strategic intelligence and individual employment decision-making is restated several

times in Sections 3–6. Moderate condensation would improve the readability of the manuscript.

Second, the description of model cards, datasheets, and algorithmic audits as effectively "minimum documentation" should be softened. The cited literature strongly supports these as useful governance artifacts, but they should perhaps be described as **recommended assurance artifacts** unless a particular legal or contractual requirement mandates them.

Third, the fairness section is informative but somewhat detached from the central classification. The discussion of demographic parity, equal opportunity, equalized odds, calibration, and the four-fifths rule could be shortened or moved to an appendix unless the author intends for fairness evaluation to constitute an independent major contribution.

Fourth, the worked example in Section 6 is useful but represents a favorable case. The framework would be much more convincing if the author added a **borderline example**, such as aggregate workforce restructuring guidance that subsequently drives decisions concerning employees in a small role family.

Finally, the operational checklist would benefit from adding explicit controls for the **intended purpose**, **reasonably foreseeable use**, **decision coupling**, and **scope-change/reclassification triggers**. These concepts appear throughout the paper but should be directly visible in the practitioner's instrument.

Declarations

Potential competing interests: No potential competing interests to declare.