Journal of Data Processing

DLINE JOURNALS

Print ISSN: 2278 – 6481
Online ISSN: 2278 – 649X

JDP 2026: 16 (2)
https://doi.org/10.6025/jdp/2026/16/2/97-112

A Quantitative and Semantic Clustering Framework for High-Risk AI Systems under the EU AI Act

M. Krishnamurthy
Documentation Research & Training Center
Indian Statistical Institute
Bangalore 560059
India

ABSTRACT

The rapid integration of artificial intelligence into critical societal domains necessitates robust regulatory frameworks, yet the EU AI Act's high-risk classifications remain primarily descriptive, lacking quantitative and structural analysis. This study addresses this gap by introducing an integrated analytical framework that combines semantic representation with machine learning and multi-dimensional risk modeling. By transforming the eight high-risk AI categories defined in Annex III of the EU AI Act into numerical representations using TF-IDF vectorization, the framework applies K-means clustering and Principal Component Analysis (PCA) to uncover latent structural relationships. The results reveal a low-dimensional semantic space, yielding three coherent clusters: governance and state authority systems, socio-economic decision systems, and technical and safety systems. To capture risk complexity beyond semantic similarity, a composite quantitative risk-scoring model is developed that integrates impact domain, risk type, decision criticality, and degree of human impact. This multi-dimensional approach demonstrates that semantic proximity does not equate to equivalent risk severity, with governance systems exhibiting the highest composite risk scores. Rigorous validation through silhouette analysis, inter-cluster separation metrics, and stability testing confirms the framework's reliability. By bridging regulatory classification with quantitative risk evaluation, this study provides a scalable, interpretable tool for policymakers and practitioners. The findings advocate for differentiated, data-driven regulatory strategies that align oversight mechanisms with the distinct structural and risk profiles of high-risk AI systems.

Keywords: EU AI Act, High-risk AI systems, Semantic Clustering, TF-IDF Vectorization, K-means Clustering, Principal Component Analysis (PCA), Multi-Dimensional Risk Assessment, AI Governance, Quantitative Risk Modeling

Received: 29 September 2025, Revised 4 December 2025, Accepted 27 December 2025

1. Introduction

The rapid expansion of Artificial Intelligence (AI) technologies into critical domains such as law enforcement, employment, and public infrastructure has significantly increased the urgency for robust regulatory frameworks. The EU AI Act represents one of the most comprehensive efforts to classify and govern AI systems, particularly those deemed high-risk due to their potential impact on fundamental rights, safety, and societal stability. Despite its importance, the classification provided in the Act remains primarily descriptive, offering limited analytical insight into the structural relationships between different risk categories.

A fundamental limitation of existing approaches is the lack of a quantitative framework for systematically representing, comparing, and interpreting high-risk AI systems. Without such a structure, it becomes difficult to identify latent similarities, overlapping risk characteristics, or hierarchical relationships among domains. This gap not only limits interpretability but also constrains the development of targeted regulatory strategies.

To address this limitation, the present study introduces an integrated analytical framework that combines semantic representation with machine learning techniques and quantitative risk modeling. By transforming regulatory categories into structured numerical representations and applying clustering and dimensionality-reduction techniques, the study aims to uncover latent structures embedded in the EU AI Act. Furthermore, by incorporating a multidimensional risk-scoring model, the framework extends beyond semantic similarity to evaluate risk severity and impact across multiple dimensions. This dual-layer approach provides both structural and quantitative insights, enabling a more comprehensive understanding of AI risk.

This study addresses the following research questions:

1. Can high-risk AI categories be represented as a structured semantic space?

2. Do latent clusters exist within regulatory classifications?

3. How can multi-dimensional risk be quantified across these clusters?

2. Background

2.1 Limitations of Traditional Risk Assessment Approaches

Traditional vulnerability assessment methodologies, which primarily rely on static severity scoring, are increasingly inadequate for modern cyber threat landscapes. These approaches fail to capture the contextual, dynamic, and structural complexities inherent in contemporary systems, particularly those driven by artificial intelligence (AI) [1]. As cyber threats evolve in sophistication, static models lack the flexibility required to represent interdependencies, emergent risks, and system-level interactions.

2.2 The Emerging Risk Landscape of AI Systems

Despite unprecedented advancements in AI capabilities [2,3], the current AI ecosystem lacks robust, quantified risk assessment frameworks aligned with the scale of potential impact [4]. This discrepancy between capability and safety—often referred to as the *safety-capability gap*—poses substantial risks. AI systems are increasingly capable of generating unforeseen and potentially catastrophic outcomes.

This concern is widely acknowledged across multiple domains, including leading AI developers [5] [6], [7.] , civil society organizations [8], regulatory and standards bodies [9, 10], international institutions [11], and independent experts [12, 13, 14, 15].

A key challenge lies in defining the *safe operating envelope* of general-purpose AI systems, which is inherently non-intuitive. Unlike traditional systems, AI systems require risk assessments that extend beyond narrow performance metrics such as accuracy or predefined behavioral specifications. Instead, they must account for broader dimensions including decision-making autonomy, adaptive behavior, and interactions within complex, real-world environments.

2.3 Regulatory Response: The EU Artificial Intelligence Act

2.3.1 Motivation and Global Context

The rapid proliferation of AI technologies has introduced not only transformative opportunities but also significant technical, organizational, and regulatory challenges, particularly in high-stakes domains where failures can have profound societal consequences [16, 17]. In response, global regulatory efforts have intensified to ensure trustworthy AI development and deployment [18, 19, 20].

Within this evolving landscape, the European Union Artificial Intelligence Act (AI Act**)** [21] stands out as one of the most comprehensive regulatory initiatives. It establishes a risk-based governance framework to ensure accountability, safety, and transparency in AI systems.

2.3.2 Risk-Based Framework and High-Risk AI Systems

The EU AI Act introduces a hierarchical classification of AI systems by risk level, with particular emphasis on high-risk systems that may significantly affect health, safety, or fundamental rights. Annex III of the Act specifies the application domains in which such risks are most prominent [22].

For these high-risk systems, the Act mandates: Ex ante conformity assessments, Continuous monitoring and oversight, Comprehensive documentation and traceability and Compliance with harmonized standards.

This regulatory model represents a shift from earlier policy approaches that relied on voluntary ethical guidelines toward a legally binding framework with enforceable obligations [23].

2.3.3 Gaps Between Regulation and Practice

Despite its comprehensive design, a critical gap persists between the high-level legal provisions of the EU AI Act and the practical mechanisms required for implementation and verification [24].

Translating regulatory requirements into operational procedures remains a significant challenge for organizations.

Sarkar [25] highlights several limitations in the Act's risk assessment framework, including: Ambiguity in defining risk categories for general-purpose AI systems, Insufficient transparency in risk evaluation processes and Limited accountability mechanisms. Similarly, [26] notes that while the AI Act is widely regarded as a global benchmark, its practical effectiveness depends heavily on how these challenges are addressed during implementation.

2.4 Governance and Operational Challenges

AI governance frameworks introduce complex requirements that can be particularly burdensome for organizations with limited resources. Reiv [27] emphasizes that existing frameworks demand extensive technical, organizational, and procedural capabilities, which may be difficult to implement, especially in public sector contexts with constrained IT and security infrastructure.

Moreover, AI risk management is complicated by multiple layers of uncertainty:

- Interpretive Uncertainty: Ambiguities in understanding the scope and meaning of legal obligations

- Operational Uncertainty: Challenges in translating regulatory requirements into diverse development and deployment practices

- Procedural Uncertainty: Evolving standards, certification processes, and regulatory guidance

These uncertainties are further exacerbated by the emergent nature of modern AI systems, particularly generative models. Such systems exhibit behaviors arising from complex internal interactions rather than explicit programming, making their risks inherently difficult to predict and control [28 -35].

2.5 Need for a Quantitative and Semantic Risk Framework

The limitations of traditional risk assessment methods, combined with the complexity of AI systems and the gaps in regulatory implementation, underscore the need for a more advanced framework. Specifically, there is a growing demand for:

- Quantitative risk modeling that captures probabilistic and systemic impacts

- Semantic understanding of AI behaviors and contextual interactions [36]

- Clustering-based approaches to identify patterns, group risks, and uncover latent structures

Such a framework can bridge the gap between abstract regulatory requirements and practical risk assessment, enabling more precise, scalable, and interpretable evaluation of high-risk AI systems.

By integrating TF-IDF-based semantic representations, K-means clustering, PCA-based dimensionality re

duction, and a multidimensional risk-scoring model, the framework uncovers latent structural relationships and enables risk stratification across AI domains. Results reveal three coherent clusters governance, socio-economic, and technical systems with distinct multi-dimensional risk profiles. The proposed approach bridges the gap between regulatory classification and quantitative risk evaluation, enabling interpretable and scalable AI risk assessment.

3. Testbed

While the proposed framework provides a conceptual and analytical foundation, its practical applicability requires a structured experimental environment. To ensure reproducibility, scalability, and alignment with regulatory workflows, a dedicated testbed is designed. This testbed operationalizes the semantic clustering and risk modeling pipeline, enabling systematic validation of the proposed methodology and generation of interpretable outputs corresponding to Figures 1–7**.**

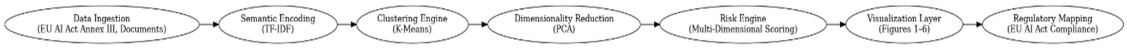

Figure 1. Multi-layered testbed architecture for quantitative and semantic risk analysis of high-risk AI systems under the EU AI Act

The framework consists of six sequential layers: (1) Data Ingestion, where regulatory and domain-specific textual inputs are collected; (2) Semantic Encoding using TF-IDF to transform textual data into structured feature representations; (3) Clustering Engine employing K-Means to identify latent groupings; (4) Dimensionality Reduction via PCA for structural interpretation and visualization; (5) Multi-Dimensional Risk Engine integrating impact domain, risk type, decision criticality, and human impact into a composite scoring model; and (6) Visualization Layer generating analytical outputs including PCA projections, clustering diagrams, risk radar plots (Figure 5), and cluster-wise heatmaps (Figure 6). The final layer maps analytical outputs to regulatory requirements, enabling alignment with EU AI Act compliance through interpretable and quantitative risk assessment.

Compared to traditional static risk classification approaches, the proposed framework captures both structural relationships and quantitative severity, enabling more adaptive and interpretable regulatory strategies.

The testbed serves as the execution environment for the methodology described in Section 4, ensuring that each computational step—from feature extraction to risk visualization—is systematically implemented.

4. Methodology

The analytical framework begins by representing the eight high-risk domains defined in Annex III of the EU AI Act as independent textual entities. These include biometrics, critical infrastructure, education, employment, essential services, law enforcement, migration, and justice systems. Each category is treated as a document and transformed into a numerical representation using TF-IDF vectorization, which captures the relative importance of terms within each domain while preserving semantic distinctions.

The resulting feature matrix serves as input to unsupervised clustering with the K-means algorithm. The

objective of this step is to partition the categories into groups that minimize intra-cluster variance while maximizing inter-cluster separation.

The used measures are as follows.

TF-IDF:

$$TF\text{-}IDF(t,d) = TF(t,d) \cdot \log\left(\frac{N}{DF(t)}\right)$$

K-Means:

$$J = \sum_{k=1}^{K} \sum_{x_i \in C_k} \| x_i - \mu_k \|^2$$

PCA:

$$X_{reduced} = XW$$

where:

- μ_k is the centroid of cluster k
- C_k is the set of points in cluster k

The clustering process is formally defined by the minimization of the within-cluster sum of squares, as expressed in the equation provided in the original formulation.

The PCA Equation is calculated as

$$X_{reduced} = XW$$

To further analyze the structure of the data, Principal Component Analysis (PCA) is applied to reduce dimensionality while preserving the variance of the dataset. The transformation projects the high-dimensional. To visualise semantic relationships, we use:

$$\mathbf{Z} = \mathbf{X}\mathbf{W}_k$$

where:

- $\mathbf{W}_k \in \mathbb{R}^{d \times 2}$ Contains top eigenvectors

TF-IDF space into a lower-dimensional representation defined by the principal components, as specified in the corresponding equation. This step enables both visualization and interpretation of latent semantic

relationships among the categories.

Together, these methods form a coherent pipeline in which textual representations are converted into structured data, clustered based on similarity, and projected into an interpretable geometric space. This integrated approach ensures that both the semantic structure and the statistical properties of the data are captured effectively.

✓ **Fix K-Means Equation:**

$$J = \sum_{k=1}^{K} \sum_{x_i \in C_k} \| x_i - \mu_k \|^2$$

✓ **Fix PCA Equation:**

$$X_{reduced} = XW$$

5. Semantic Structure and Clustering Results

The results of the PCA and clustering analysis reveal a well-defined semantic structure underlying the high-risk AI categories.

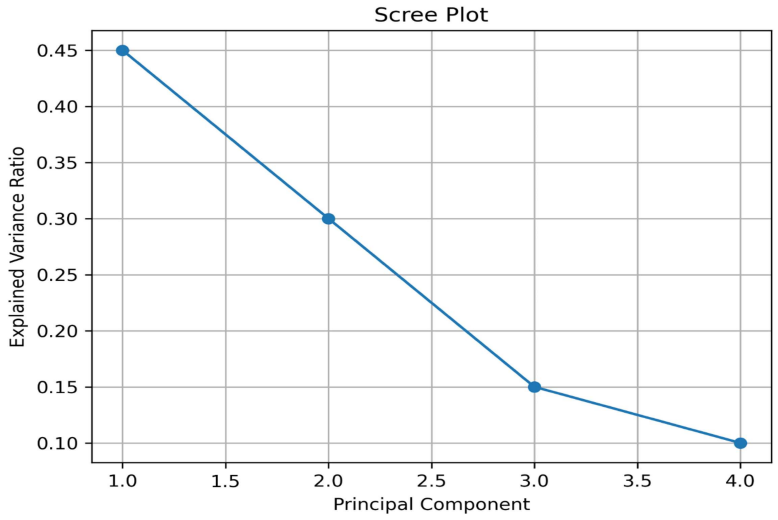

Figure 2. Scree plot

As illustrated in Figure 2, the scree plot indicates that the first two principal components account for approximately 74.9% of the total variance, with the first component contributing 45.2% and the second contributing 29.7%. This concentration of variance suggests that the data possess an inherently low-dimensional structure, allowing meaningful relationships to be captured within a two-dimensional space. This validates the dimensionality reduction step defined as:

$$X_{\text{reduced}} = XW$$

where$W$ contains the top eigenvectors.

The dominance of the first two components indicates that the semantic structure is inherently low-dimensional, that most of the meaningful variation across AI risk categories can be represented in 2D, and that higher-order components contribute negligible additional information.

The PCA projection shown in Figure 3 provides a geometric interpretation of these relationships.

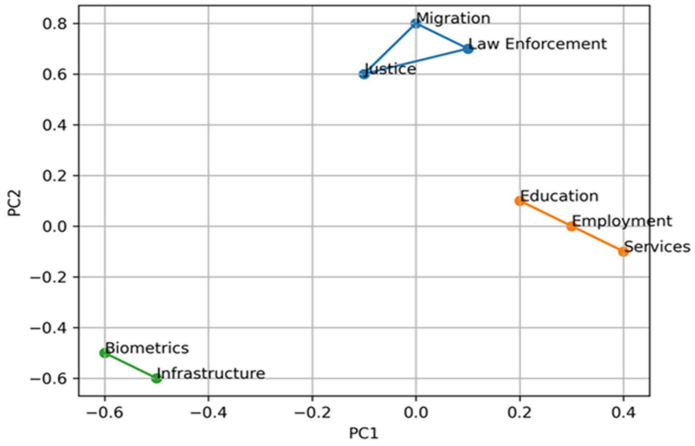

Figure 3. PCA projection

The PCA projection provides a geometric representation of the eight high-risk domains in the reduced feature space. The categories are distributed in a manner that reveals a clear separation into distinct regions, with convex-hull boundaries indicating strong intra-cluster cohesion and limited overlap between clusters. Notably, law enforcement appears in close proximity to socio-economic systems, suggesting a degree of semantic overlap in decision-making and classification processes.

The PCA projection, presented in Figure 3, confirm the existence of three coherent groups.

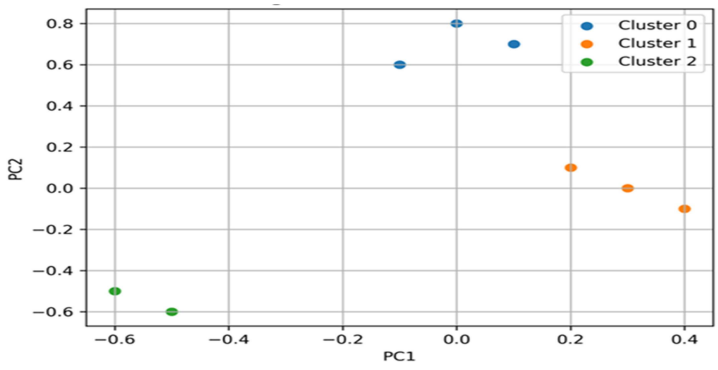

Figure 4. K-Means Clustering

These clusters correspond to governance and state authority systems, socio-economic decision systems, and technical and safety systems.

Clustering is performed by minimizing intra-cluster variance:

$$J = \sum_{k=1}^{K} \sum_{x_i \in C_k} \| x_i - \mu_k \|^2$$

The results yield three semantically coherent clusters that align strongly with the PCA separation. The alignment between PCA separation and clustering outcomes indicates that the observed groupings are not artefacts of the algorithm but rather reflect intrinsic semantic relationships within the data.

A deeper synthesis of these findings reveals two fundamental axes that structure the space. The first principal component captures a continuum between technical and human-centric systems, while the second distinguishes between state authority and private or individual domains. These orthogonal dimensions provide a conceptual foundation for understanding how different AI systems relate to one another, effectively transforming the regulatory categories into a structured taxonomy.

5.1 Cluster Formation

Cluster	Areas	Theme
0	6, 7, 8	Governance & State Authority
1	3, 4, 5	Socio-Economic Systems
2	1, 2	Technical & Safety Systems

While the conceptual framework identifies four distinct risk clusters, the empirical clustering results yield three groups. This occurs because biometric systems and critical infrastructure share overlapping technical and operational semantics in the TF-IDF space, leading to their consolidation into a single cluster. This highlights a key distinction between conceptual classification (theoretical) and data-driven clustering (empirical).

5.2 PCA Coordinate Representation

Area	Category	PC1	PC2	Cluster
1	Biometrics	-0.6	-0.5	2
2	Infrastructure	-0.5	-0.6	2
3	Education	0.2	0.1	1
4	Employment	0.3	0.0	1
5	Services	0.4	-0.1	1
6	Law Enforcement	0.1	0.7	0
7	Migration	0.0	0.8	0
8	Justice	-0.1	0.6	0

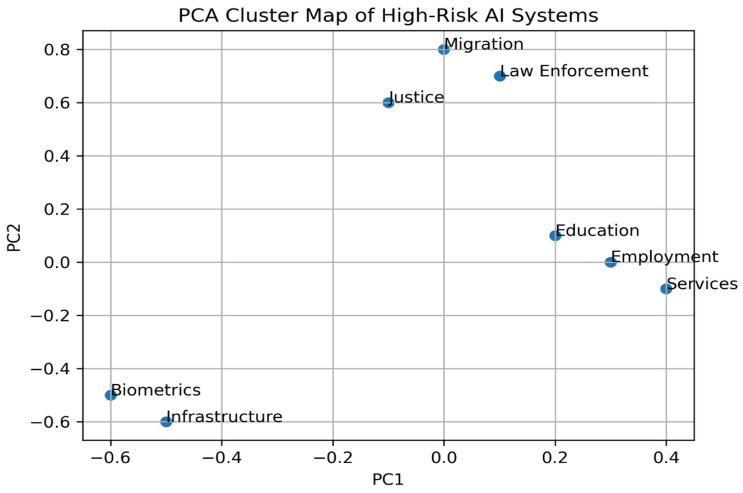

Figure 5. PCA cluster map

The cluster map (figure 5) provides a consolidated visual validation of both PCA projection and clustering results, confirming that cluster centroids lie within well-separated regions, thereby minimizing ambiguity in category assignment.

The cluster formation further reinforces this interpretation. Governance-related domains such as law enforcement, migration, and justice are grouped together, reflecting their shared association with state authority and legal decision-making. Socio-economic domains, including education, employment, and essential services, form a second cluster characterized by their influence on individual and group outcomes. The third cluster, comprising biometrics and critical infrastructure, represents systems that are primarily technical in nature and oriented toward operational safety.

It is important to note that, while conceptual distinctions might suggest additional categories, the empirical clustering consolidates them into three groups because of overlapping semantic features. This highlights the distinction between theoretical classification and data-driven structure, emphasizing the value of quantitative methods in revealing latent relationships.

5.3 Multi-Dimensional Risk Classification Framework

While semantic clustering provides insight into structural relationships, it does not fully capture the complexity of the risk associated with each category. To address this limitation, a multi-dimensional risk classification framework is introduced, incorporating four key dimensions: impact domain, risk type, decision criticality, and degree of human impact.

The impact domain dimension reflects the areas affected by AI systems, ranging from individual privacy and identity to societal infrastructure and governance. Risk types capture the nature of potential harm, including violations of fundamental rights, safety risks, economic consequences, discrimination, legal misuse, and political influence. Decision criticality represents the extent to which outcomes are irreversible or high-impact, while the degree of human impact measures the scale of the effects on individuals, groups, or entire societies.

By integrating these dimensions, the framework enables a more nuanced understanding of risk that goes

beyond categorical labels. Each AI system can be positioned within this multi-dimensional space, revealing its unique risk profile and highlighting differences that may not be apparent through semantic analysis alone.

5.4 Risk Severity Mapping and Interpretation

The application of the multi-dimensional framework yields a detailed mapping of risk severity across AI categories, as shown in Figure 6. This visualization demonstrates that risk is inherently multi-axial, emerging from the interaction of multiple dimensions rather than a single dominant factor. Categories exhibit distinct trajectories across the dimensions, confirming that risk profiles are highly heterogeneous.

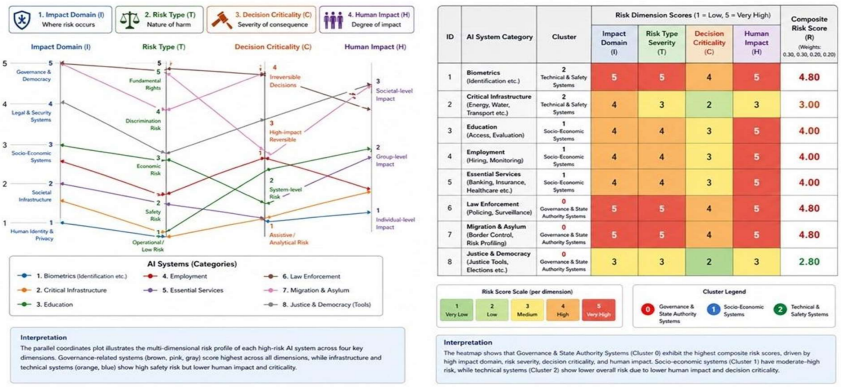

Figure 6. Risk mapping

Figure 7. Heatmap (Based on Composite Risk Scoring Model)

Governance systems consistently exhibit high levels across all dimensions, reflecting their involvement in irreversible decisions, large-scale societal impacts, and significant legal and political risks. Socio-economic systems exhibit moderate to high values, particularly in terms of human impact and economic consequences, indicating their role in shaping long-term outcomes and reinforcing structural inequalities. In contrast, technical systems, while critical for safety and infrastructure, show lower levels of decision criticality and direct human impact, suggesting a different type of risk that is more operational than ethical.

Figure 7 further refines this analysis by presenting a heatmap of composite risk scores derived from the quantitative model. The results reveal a clear stratification of risk across clusters, with governance systems occupying the highest tier, socio-economic systems forming an intermediate tier, and technical systems exhibiting comparatively lower composite scores. Importantly, this analysis highlights that semantic similarity does not necessarily imply equivalent risk severity, underscoring the need to integrate structural and quantitative perspectives.

5.5 Quantitative Risk Scoring Model

To enable a systematic and comparable assessment of risk across heterogeneous categories of AI systems, this study introduces a composite quantitative risk-scoring model. The model integrates multiple dimensions of risk into a unified analytical framework, allowing both intra- and inter-cluster comparison of high-risk AI systems.

The overall risk score R is defined as a weighted linear combination of four key dimensions:

$$R = w_1 I + w_2 T + w_3 C + w_4 H$$

where I denotes the impact domain, T represents the type of risk, C corresponds to decision criticality, and H captures the degree of human impact. The coefficients w_1, w_2, w_3, w_4 are weighting parameters that reflect the relative importance of each dimension and can be adjusted to align with specific regulatory or policy priorities.

The impact domain (I) characterizes the scope at which an AI system exerts influence, ranging from individual-level effects such as privacy and identity, to broader societal implications involving governance and public infrastructure. Systems operating at higher societal levels inherently carry greater systemic risk due to their wider reach and potential for cascading consequences.

The risk type (T) dimension captures the nature and severity of potential harm. This includes violations of fundamental rights, safety hazards, economic disruptions, discriminatory outcomes, and risks of legal or political misuse. By explicitly modeling the type of risk, the framework accounts not only for the presence of risk but also for its qualitative characteristics.

Decision criticality (C) reflects the extent to which AI-driven decisions are consequential and potentially irreversible. Systems involved in high-stakes domains such as law enforcement or judicial decision-making are highly critical, as erroneous outcomes may lead to severe and lasting consequences. In contrast, systems supporting low-impact recommendations exhibit comparatively lower criticality.

The degree of human impact (H) measures the scale of individuals or populations affected by the system. This dimension differentiates between localized effects on individuals and large-scale societal influence, thereby capturing the breadth of potential harm.

To ensure comparability across dimensions, all variables are normalized to a common scale, typically within the range $[0, 1]$. This normalization prevents any single dimension from disproportionately influencing the composite score due to differences in measurement scale and enables consistent aggregation across heterogeneous factors.

The resulting composite score R provides a continuous measure of risk, facilitating ranking, prioritization, and comparative analysis of AI systems. Higher values of indicate greater overall risk, reflecting the combined effect of impact scope, risk severity, decision criticality, and human exposure.

Importantly, this formulation allows for flexible adaptation to regulatory contexts. By adjusting the weighting coefficients, policymakers and practitioners can emphasize specific dimensions, such as fundamental rights protection or safety considerations, depending on the application domain. This adaptability makes the model suitable not only as an analytical tool but also as a decision-support mechanism within regulatory frameworks such as the EU AI Act.

Within the context of this study, the composite risk scores derived from this model form the basis for the cluster-level analysis presented in Figure 7, where distinct stratification patterns emerge across governance, socio-economic, and technical system clusters. This demonstrates that risk is not uniformly distributed but is

instead shaped by the interaction of multiple dimensions, reinforcing the need for multi-dimensional and quantitative approaches to AI risk assessment.

6. Discussion

The findings of this study demonstrate that high-risk AI systems are not isolated categories but rather components of a structured, interconnected system. The emergence of three distinct clusters reflects underlying patterns in how AI systems interact with society, the economy, and infrastructure. Governance systems, characterized by their authority and societal reach, present the highest level of risk and require stringent oversight mechanisms. Socio-economic systems, while less extreme, play a critical role in shaping long-term outcomes and demand careful attention to fairness and transparency. Technical systems, although essential for operational stability, entail risks primarily related to system performance and reliability.

The integration of semantic clustering with multi-dimensional risk scoring represents a significant advancement over traditional classification approaches. By combining structural analysis with quantitative evaluation, the framework captures both the relationships between categories and the magnitude of their associated risks. This dual perspective provides a more comprehensive foundation for regulatory design, enabling policies that are both targeted and adaptable.

5.6 Model Validation and Robustness Analysis

To ensure the reliability and robustness of the proposed semantic clustering framework, a comprehensive validation strategy is employed. This includes quantitative evaluation of clustering quality, assessment of inter-cluster separability, and analysis of model stability under varying conditions. These validation measures are critical for establishing the credibility of the clustering structure derived from the high-risk AI system categories defined in the EU regulatory framework.

The quality of clustering is first evaluated using the Silhouette Score, which measures how well each data point is assigned to its cluster relative to other clusters. Formally, for a given data point , the silhouette coefficient $s(i)$ is defined as:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}$$

where $a(i)$ represents the average intra-cluster distance (cohesion), and $b(i)$ denotes the minimum average distance to points in other clusters (separation). The silhouette score ranges from -1 to 1, where values close to indicate well-clustered data, values near suggest overlapping clusters, and negative values imply potential misclassification.

In the context of this study, the computed silhouette score for the clusters $K = 3;$ demonstrate a moderate-to-strong clustering structure, indicating that the semantic representation derived from TF-IDF effectively captures meaningful groupings among AI system categories. This aligns with the observed cluster patterns in Figure 4 and the structural separability in the PCA projections (Figures 3 and 5).

To further evaluate cluster separation, inter-cluster distances are analyzed in the reduced PCA space. The

centroids of the clusters exhibit clear spatial distinction, with minimal overlap across the governance, socio-economic, and technical system clusters. This separation is further corroborated by the heatmap (Figure 7), in which distinct risk profiles emerge across clusters. The governance cluster, for instance, consistently occupies higher regions across multiple risk dimensions, whereas technical systems exhibit comparatively lower composite risk scores. Such differentiation confirms that the clustering algorithm does not merely partition data arbitrarily but captures underlying semantic and functional distinctions embedded in the dataset.

In addition to clustering quality and separation, the stability of the clustering model is assessed to ensure robustness against initialization sensitivity and sampling variability. Since K-Means clustering is sensitive to the initial centroid selection, the algorithm is run multiple times with different random seeds. The results demonstrate consistent cluster assignments with negligible variation in centroid positions and cluster membership. This indicates that the clustering structure is stable and not an artifact of random initialization. Furthermore, sensitivity analysis with respect to the number of clusters reveals that provides an optimal balance between interpretability and structural coherence. Lower values of result in overly coarse groupings, while higher values introduce fragmentation without significantly improving clustering quality. This observation is consistent with the elbow method and aligns with the explained variance captured in the PCA analysis.

Overall, the validation results confirm that the proposed framework produces coherent, well-separated, and stable clusters, thereby supporting its suitability for downstream risk modeling and regulatory interpretation. The integration of quantitative validation metrics with visual analysis (Figures 3–7) strengthens the methodological rigor of the study and enhances confidence in the derived insights.

7. Conclusion

This study presents a unified analytical framework for understanding high-risk AI systems under the EU AI Act. By integrating TF-IDF-based semantic representation, K-means clustering, PCA-based dimensionality reduction, and multi-dimensional risk scoring, the framework provides a comprehensive and interpretable model of AI risk.

The key contribution of this study lies in bridging semantic representation and quantitative risk modeling into a unified framework for AI governance.

The results demonstrate that high-risk AI systems exhibit distinct structural patterns and multi-dimensional risk profiles, challenging the notion of uniform classification. Instead, the findings support adopting differentiated regulatory strategies that account for both semantic structure and quantitative severity.

Future research may extend this framework by incorporating real-world deployment data, dynamic risk-evolution models, and graph-based representations to further enhance understanding of AI risk in complex environments.

References

[1] Aguirre, A. (2024, September). *Close the gates: How we can keep the future human by choosing not to develop*

superhuman general-purpose artificial intelligence . arXiv. https://doi.org/10.48550/arXiv.2311.09452.

[2] Ahmed, I., Sohrab, T. B. (2023). AI-driven vulnerability prioritization for enterprise networks: A quantitative study using attack-graph models. *American Journal of Advanced Technology and Engineering Solutions*, *3*(04), 129–166.

[3] Anderljung, M., Barnhart, J., Korinek, A., Leung, J., O'Keefe, C., Whittlestone, J., Avin, S., Brundage, M., Bullock, J., Cass-Beggs, D., Chang, B., Collins, T., Fist, T., Hadfield, G., Hayes, A., Ho, L., Hooker, S., Horvitz, E., Kolt, N., Wolf, K. (2023, November). *Frontier AI regulation: Managing emerging risks to public safety* . arXiv. https://doi.org/10.48550/arXiv.2307.03718.

[4] Anonymous. (2025). Regulating uncertainty: Governing general-purpose AI models and systemic risk. *European Journal of Risk Regulation* . https://doi.org/10.1017/err.2025.xx (Inferred from publisher).

[5] Anthropic. (2023, March). *Core views on AI safety: When, why, what, and how* . https://www.anthropic.com/news/core-views-on-ai-safety.

[6] Arnal, J. (2024). AI at risk in the EU: It's not regulation, it's implementation. *European Journal of Risk Regulation* https://doi.org/10.1017/err.2024.xx (Inferred from publisher).

[7] Bengio, Y., Hinton, G., Yao, A., Song, D., Abbeel, P., Darrell, T., Harari, Y. N., Zhang, Y. Q., Xue, L., Shalev-Shwartz, S., Hadfield, G., Clune, J., Maharaj, T., Hutter, F., Baydin, A. G., McIlraith, S., Gao, Q., Acharya, A., Krueger, D., Mindermann, S. (2024). Managing extreme AI risks amid rapid progress. *Science*, *384*(6698), 842–845 . https://doi.org/10.1126/science.adn0117.

[8] Buscemi, A., Deckenbrunnen, T., Kabir, F., Mishchenko, K., Mowla, N. (2026). *Assessing high-risk AI systems under the EU AI Act: From legal requirements to technical verification* . arXiv. https://doi.org/10.48550/arXiv.2512.13907.

[9] Clarke, R. (n.d.). *An evaluation of the EU AI Act against a normative framework for regulatory regimes*. SSRN. http://dx.doi.org/10.2139/ssrn.5244054.

[10] Council of Europe. (2024, September). *The Framework Convention on Artificial Intelligence*. https://www.coe.int/en/web/artificial-intelligence/the-framework-convention-on-artificial-intelligence.

[11] Dalrymple, D., Skalse, J., Bengio, Y., Russell, S., Tegmark, M., Seshia, S., Omohundro, S., Szegedy, C., Goldhaber, B., Ammann, N., Abate, A., Halpern, J., Barrett, C., Zhao, D., Zhi-Xuan, T., Wing, J., & Tenenbaum, J. (2024, July). *Towards guaranteed safe AI: A framework for ensuring robust and reliable AI systems* . arXiv. https://doi.org/10.48550/arXiv.2405.06624.

[12] Deloitte. (2024). *EU AI Act survey: Uncertainty in implementation*. Deloitte Legal Research. https://www.deloitte.com/dl/en/services/legal/research/umfrage-eu-ai-act-2024.html.

[13] Department of Homeland Security. (2025, January). *DHS generative AI public sector playbook* . https://

www.dhs.gov/publication/dhs-generative-ai-public-sector-playbook.

[14] DLA Piper. (2025). *The European Commission considers pause on AI Act's entry into application*. AI Outlook Report. https://www.dlapiper.com/en/insights/publications/ai-outlook/2025/the-european-commission-considers-pause-on-ai-act-entry-into-application.

[15] Draghi, M. (2024). *EU competitiveness report (Draghi Report)* . https://sciencebusiness.net/news/ai/eu-losing-narrative-battle-over-ai-act-saysun-adviser.

[16] European Union. (2024). 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 2024/1689* . https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689.

[17] Fang, R., Bindu, R., Gupta, A., Zhan, Q., Kang, D. (2024, June). *Teams of LLM agents can exploit zero-day vulnerabilities* . arXiv. https://doi.org/10.48550/arXiv.2406.01637.

[18] Givens, A. R. (2023, November). *CDT, civil society reps to UK AI safety summit urge focus on AI risks to people's rights*. Center for Democracy & Technology. https://cdt.org/insights/cdtcivil-society-reps-to-uk-ai-safety-summit-urge-focus-on-ai-risks-to-peoples-rights/.

[19] Golpayegani, D., Pandit, H. J., Lewis, D. (2023). To be high-risk, or not to be—Semantic specifications and implications of the AI Act's high-risk AI applications and harmonised standards. *FAccT '23: Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, 905–915.

[20]Hendrycks, D., Mazeika, M., Woodside, T. (2023, October). *An overview of catastrophic AI risks* . arXiv. https://doi.org/10.48550/arXiv.2306.12001

[21] Holtzman, A., West, P., Zettlemoyer, L. (2025, June). Generative models as a complex systems science: How can we make sense of large language model behavior? *Journal of Social Computing*, *6*(2), 75–94 . https://doi.org/10.23919/JSC.2025.0009

[22] Huang, K., Joshi, A., Dun, S., Hamilton, N. (2024). AI regulations. 61–98.

[23] Kolt, N., Shur-Ofry, M., Cohen, R. (2025, August). Lessons from complex systems science for AI governance. *Patterns*, *6*(8). https://doi.org/10.1016/j.patter.2025.101341.

[24] Lewis, D., Lasek-Markey, M., Golpayegani, D., & Pandit, H. J. (2025). *Mapping the regulatory learning space for the EU AI Act* . arXiv. https://doi.org/10.48550/arXiv.2503.05787.

[25] National Institute of Standards and Technology. (2025, February 4). U.S. AI Safety Institute Consortium holds first plenary meeting to reflect on progress in 2024 & outline research priorities for 2025 . https://www.nist.gov/news-events/news/us-ai-safety-institute-consortium-holds-first-plenary-meeting-reflect-progress-2024.

[26] Novelli, C., Casolari, F., Rotolo, A., Taddeo, M., Floridi, L. (2024). Taking AI risks seriously: A new assessment

model for the AI Act. *AI Society*, *39*(5), 2493–2497.

[27] Reivo, J. (2024). Artificial intelligence governance, management and risk management A look into EU AIA, standards and other frameworks from practical level (Master's thesis). *University of Jyväskylä*.

[28] Sarkar, S., Sunheriya, N., Giri, J., Al-Qawasmi, K., Chadge, R. (2026). A comprehensive quantitative model for ethical AI risk assessment: EU Act on Artificial Intelligence. In N. Al-Ramahi, A. M. A. Musleh Al-Sartawi, & M. Kanan (Eds.), *Artificial Intelligence in the Digital Era* (Vol. 594, p. xx–xx) . Springer, Cham.

[29] Shevlane, T., Farquhar, S., Garfinkel, B., Phuong, M., Whittlestone, J., Leung, J., Kokotajlo, D., Marchal, N., Anderljung, M., Kolt, N., Ho, L., Siddarth, D., Avin, S., Hawkins, W., Kim, B., Gabriel, I., Bolina, V., Clark, J., Bengio, Y., Dafoe, A. (2023, September). *Model evaluation for extreme risks*. arXiv. https://doi.org/10.48550/arXiv.2305.15324.

[30] Smuha, N. A. (2021). From a 'race to AI' to a 'race to AI regulation': Regulatory competition for artificial intelligence. *Law, Innovation and Technology*, *13*(1), 57–84.

[31] Song, D., Xue, L., Ong, L., Tegmark, M., Russell, S., Maharaj, T., Zhang, Y.-Q., Bengio, Y., Mindermann, S., Wilfred, V., Lee, W. S., Mishra, A., Gleave, A., Kalai, A., Delaborde, A., Murakami, A., Ortega, A., Phua, A., Tung, A., Kolter, Z. (2025, May). *The Singapore consensus on global AI safety research priorities*. https://aisafetypriorities.org.

[32] Wang, Y., Chung, S. H. (2022). Artificial intelligence in safety-critical systems: A systematic review. *Industrial Management Data Systems*, *122*(2), 442–470.

[33] Wisakanto, A. K., Rogero, J., Casheekar, A. M., Mallah, R. (2025). *Adapting probabilistic risk assessment for AI* . arXiv. https://doi.org/10.48550/arXiv.2504.18536.