Profiling Cross-Border Remote Cybersecurity Employment for Jurisdictional Complexity via Unsupervised Role-Arrangement Clustering

Ahmed Saeed Bahurmuz[1,*], Mohamed Ahmed Alhebi[2]

[1,2]Information Science Department, King Abdulaziz University, Jeddah, Saudi Arabia

ABSTRACT

Cross-border remote work has expanded the practical footprint of cybersecurity labor beyond traditional organizational and national boundaries, raising jurisdictional questions about employment governance, taxation, and cross-border data access. However, empirical cyberlaw research is often limited by the availability of legal outcome variables at scale. This study presents a dataset-only, unsupervised method to detect and summarize cross-border remote work configurations that plausibly differ in jurisdictional coordination burden. Using a structured salary dataset of cybersecurity-related roles, we operationalize cross-border status as a mismatch between employee residence and company location. To avoid a trivial domestic-versus-cross-border split, clustering is performed exclusively on cross-border records. Mixed categorical and numeric features—remote-work intensity, employment type, experience level, company size, and optional role grouping and jurisdiction categories—are represented via Gower distance. We apply average-linkage hierarchical clustering to the resulting precomputed distance matrix and select a solution using silhouette score. To prevent unstable singleton patterns from being reported as typologies, we enforce a minimum cluster size (n≥5) by merging micro-clusters into the nearest larger cluster using mean inter-cluster Gower distance. The final cross-border typology yields two interpretable groups: a dominant "enterprise remote-first" configuration (n=52) with high fully remote prevalence and strong concentration in large firms, and a smaller "mid/small mixed-remote" configuration (n=6) with no large-firm representation and a tighter salary distribution. The dominant group spans many distinct residence→company corridors, suggesting broader cross-jurisdiction exposure and coordination needs, while the smaller group reflects more constrained organizational settings. The study contributes an interpretable clustering workflow with explicit micro-cluster handling for exploratory cyberlaw analyses, and it delineates the limits of inference when legal compliance variables are absent.

Keywords cross-border remote work; jurisdictional complexity; unsupervised clustering; Gower distance; cybersecurity labor markets

Submitted 16 October 2025
Accepted 20 November 2025
Published 1 December 2025

Corresponding author
Ahmad Saeed Bahurmuz,
abahurmuz@stu.kau.edu.sa

Additional Information and Declarations can be found on page 356

DOI: 10.63913/jcl.v1i4.45

Introduction

Remote work has accelerated transnational hiring across technology sectors, including cybersecurity, reshaping role distribution, compensation practices, and where work is effectively performed [1]. Broader remote-work scholarship likewise emphasizes how globalization of employment increasingly outpaces national legal and institutional frameworks, creating new frictions in cross-border work governance [2]. Cybersecurity is a particularly sensitive case because many security practitioners operate with privileged access to critical infrastructure, production environments, and incident-response data—conditions that intensify cross-border concerns around data sovereignty, regulatory compliance in cloud environments, and access-control enforcement

How to cite this article: A.S. Bahurmuz and M.A. Alhebi, "Profiling Cross-Border Remote Cybersecurity Employment for Jurisdictional Complexity via Unsupervised Role-Arrangement Clustering," *J. Cyber. Law.*, vol. 1, no. 4, pp. 344-358, 2025.

compared with many other remote occupations [3],[4],[5].

From a cyberlaw standpoint, cross-border remote work raises “jurisdictional complexity” because traditional connecting factors for applicable employment regulation—such as the place where work is performed (lex loci laboris) versus the employer’s location—become ambiguous when work is delivered remotely across borders [6]. This ambiguity can shift or multiply legal nexuses relevant to labor protections, social security coordination, and related obligations for both employers and workers [6]. In addition, cross-border cybersecurity work frequently intersects with data governance: remote access may involve personal data, regulated operational data, or security telemetry subject to different privacy and sectoral regimes, increasing the operational burden of aligning policies and controls across jurisdictions [6], [7]. Importantly, these burdens arise even absent misconduct: legal uncertainty can elevate transaction costs, complicate supervision and accountability, and motivate divergent national responses, strengthening calls for coordination mechanisms that can handle cross-border digital work at scale [2], [8].

Despite growing attention to remote work and global labor markets, empirical cyberlaw research faces a practical constraint: most available workforce datasets contain rich occupational and compensation variables but lack direct legal outcome variables (e.g., compliance determinations, dispute outcomes, or formal jurisdiction tests). As a result, the literature offers limited data-driven typologies of cross-border remote arrangements that can serve as empirical anchors for discussing jurisdictional complexity. There is a need for an interpretable, dataset-only approach that can still align with cyberlaw questions by identifying recurring configurations associated with higher coordination burden, without over-claiming legal conclusions.

This study addresses that gap by developing an unsupervised typology of cross-border cybersecurity work arrangements from a single structured dataset. The objective is to detect and characterize cross-border remote work patterns that plausibly differ in jurisdictional coordination burden. Specifically, we ask: RQ1) What typologies of cross-border cybersecurity work exist in the dataset? RQ2) How do these typologies differ in remote intensity, firm scale, and corridor diversity (employee residence → company location pairs)? RQ3) How do compensation distributions differ across the discovered typologies?

To answer these questions, we cluster only cross-border records to avoid a trivial domestic-versus-cross-border split. We represent mixed numeric and categorical features using Gower distance and perform average-linkage hierarchical clustering with internal validation via silhouette score. To prevent over-interpretation of rare patterns, we enforce a minimum cluster size rule, merging micro-clusters into the nearest larger cluster based on mean inter-cluster Gower distance (with an outlier labeling option as a robustness alternative).

This work makes three contributions. Methodologically, it provides an interpretable unsupervised pipeline for mixed data with explicit micro-cluster handling. Empirically, it offers a cross-border typology of cybersecurity work arrangements with corresponding compensation dispersion patterns. Substantively for cyberlaw, it proposes a measurable, dataset-only entry point for analyzing jurisdictional complexity in cross-border remote work, suitable for extension when richer legal and governance variables become available.

Literature Review

Cross-border remote work and global labor markets

Remote work’s role as an enabling infrastructure for international hiring builds on long-standing telecommuting trends and intensified sharply during and after the COVID-19 period due to rapid digital adoption and continuity needs [9], [10]. As connectivity and collaboration platforms reduce coordination costs, employers can increasingly source labor beyond local markets, expanding the feasible “hiring radius” across borders [9], [10]. Empirical accounts of telework adoption highlight how organizations reconfigure processes (platform adoption, digital workflows, distributed coordination) to sustain productivity, while simultaneously introducing new managerial and infrastructural demands [10]. These shifts create opportunities to access broader talent pools, but also introduce cross-border complexity because employment relationships now traverse multiple regulatory and institutional environments [10].

Compensation and labor mobility under cross-border remote work face competing pressures. Employers may try to arbitrage geographic wage differentials, but must also manage retention, equity, and competitiveness across distributed staff [9]. Worker outcomes research emphasizes heterogeneity in preferences and constraints (e.g., home-office resources, boundary management, work–life balance) that interact with compensation and non-wage conditions in remote arrangements [11], [12]. A central challenge is the “where work happens” ambiguity: telework decouples residence from a physical workplace, complicating the designation of a single place of work for governance and regulatory purposes [11]. As a result, organizational policy choices—such as formal location designations and remote-work agreements—become key instruments for reducing ambiguity and managing downstream legal and administrative obligations [10].

Cybersecurity work as a distinct remote-work domain

Cybersecurity is a distinctive remote-work domain because it often requires elevated trust, privileged access, and tightly governed incident-response capabilities. Research on workforce information systems and compliance stresses that HR and operational systems process sensitive data, demanding strong access controls, role-based privileges, and audit-ready governance—requirements that become more acute in distributed settings [13]. Pandemic-era evidence also documents increased cyber threats (e.g., phishing and fraud campaigns) that exploited rapid transitions to home-working environments, raising the security burden on organizations and remote workers alike [14], [15]. In this context, secure remote access, privileged credential management, and controlled incident-response workflows become foundational for cybersecurity roles performed at a distance [13].

Cross-border security operations frequently involve distributed monitoring and follow-the-sun practices to achieve continuous coverage, which can improve responsiveness but also intensifies coordination demands across time zones and legal domains [10]. To mitigate risks from dispersed privileged access, organizations increasingly adopt Zero Trust concepts—continuous verification and least privilege—implemented through granular access controls and monitoring [16], [17]. The literature emphasizes that such technical architectures are most effective when paired with organizational change: policy alignment, training, incident playbooks, and governance mechanisms that scale to distributed operations [17]. Consequently, remote cybersecurity work elevates the importance of integrated technical and organizational controls as prerequisites for safe cross-border practice [13], [16], [17].

Cyberlaw and jurisdiction: frameworks relevant to cross-border work

Legal governance of cross-border work depends on multiple jurisdictional anchors—worker residence, employer location, place of work/services performed, and establishment—rather than a single determinant [18], [19]. Policy scholarship on co-occurring legal regimes argues that interacting rules should be analyzed jointly, because single-regime analyses can misrepresent how legal environments operate in practice [18]. This insight matters for remote work because the relevant connecting factors can shift with living arrangements, travel, employer structuring, and contractual definitions, producing overlapping obligations. Accordingly, cyberlaw-relevant inquiry often needs to treat jurisdiction as multi-dimensional and operationalize it through observable proxies while acknowledging that definitive determinations remain fact-dependent [18].

Employment governance in cross-border remote arrangements therefore hinges on how contracts and organizational policies interact with statutory obligations across jurisdictions. Telework research highlights that organizational policy frequently shapes day-to-day remote work, but cannot be separated from employment standards, social protections, and definitional rules that may apply by virtue of residence or the legally defined workplace [9], [12]. Compliance frameworks for workforce and HR systems likewise stress that employer governance choices (contracting terms, payroll practices, benefits administration, access policies) must incorporate both operational needs and applicable legal obligations in the worker's jurisdiction(s) [13]. The same is true for data governance: remote access and cross-border data flows intersect with privacy and transfer constraints, motivating the integration of contractual clauses, data minimization, technical controls, and continuous monitoring into organizational policy [13], [17], [20].

Empirical measurement of "jurisdictional complexity"

Operationalizing "jurisdictional complexity" for empirical research requires translating a multi-dimensional legal concept into measurable indicators. Policy-clustering and comparative multi-scheme research demonstrates approaches that code the presence and co-occurrence of relevant rules and then apply clustering or dimension reduction to characterize regimes and typologies [18], [19], [21]. Across domains, complexity is often represented by combinations of categorical indicators (e.g., policy features), continuous measures (e.g., intensity or frequency of cross-border ties), and structural variables (e.g., organizational scale) that jointly approximate the underlying construct [21], [19]. This motivates empirical designs that combine multiple proxies rather than relying on any single variable to represent jurisdictional complexity.

In practice, common proxies include remote-work intensity, contracting modality (employee versus contractor), firm size or multinational footprint, and the presence of formal organizational policies that shape cross-border work [10], [13], [22], [23]. Cluster and latent-class studies in adjacent areas (e.g., commuting, travel, and consumer profiling) illustrate how socio-demographic and attitudinal covariates can be used to interpret segments and validate whether clusters correspond to meaningful behavioral patterns [22], [24], [23]. However, methodological work cautions that proxy-based inference is limited when legal outcomes (e.g., determinations of applicable law, tax liability, employment status disputes) are unavailable; therefore, studies relying on proxies should emphasize interpretive restraint, triangulation, and sensitivity analyses to avoid treating typologies as legal determinations [18], [21].

Machine learning in legal and cyberlaw research

Machine learning supports multiple methodological aims in legal and cyberlaw research. Supervised learning is commonly used for prediction tasks when labeled outcomes exist (e.g., classification or forecasting), enabling evaluation via standard performance metrics [18], [23]. Unsupervised learning—particularly clustering—serves exploratory goals such as typology discovery, segmentation, and identifying latent structure in multi-dimensional policy or organizational data when outcomes are not directly observed [19], [21], [22]. This distinction is central in empirical cyberlaw contexts, where reliable labels for legal outcomes are often unavailable at scale, making unsupervised approaches attractive for structuring analysis and generating hypotheses.

Clustering is especially valuable for policy typologies and risk profiling because it can group observations by shared configurations of features and then support comparative interpretation using covariates (e.g., organizational scale, work arrangements, corridor diversity) [19], [22], [23]. Across policy and applied settings, scholars emphasize that interpretability is a requirement rather than a convenience: typologies intended to inform governance must be transparent, auditable, and accompanied by robustness checks [18], [13]. Compliance-oriented frameworks similarly argue that explainability is necessary for auditability and defensibility, implying that interpretable clustering workflows—with clear handling of rare cases and explicit sensitivity analyses—are preferable in cyberlaw-adjacent applications [13], [18].

Mixed-type clustering methods for socio-legal datasets

Socio-legal datasets are typically mixed-type, combining numeric measures (e.g., intensities, counts, magnitudes) with categorical indicators (e.g., contract type, jurisdiction categories, policy presence) and ordinal covariates (e.g., levels or ratings) [21],[22],[23],[24]. The literature shows multiple strategies: latent class/profile models for categorical-heavy data; numeric-only clustering after dimension reduction; and categorical PCA or related transformations to integrate categorical and continuous information into a common representational space [21],[22],[23]. These approaches share a practical goal: producing stable, interpretable clusters that can be described and compared in ways that stakeholders can understand and evaluate.

Best practices emphasize algorithm–data alignment and validation. Researchers recommend selecting clustering methods suited to the dominant data types and reporting internal validation metrics cautiously, because such metrics cannot guarantee substantive validity—especially when downstream legal interpretation is at stake [18], [23]. Accordingly, studies aimed at policy relevance often combine internal validation with robustness checks across algorithms and parameter settings, and interpret clusters using external covariates or expert review where possible [19], [23]. This motivates workflows that (i) preserve interpretability, (ii) explicitly manage outliers and micro-clusters, and (iii) document sensitivity to key parameter choices to communicate uncertainty transparently [18], [23].

Synthesis and practical implications for this study

Taken together, the literature indicates that cross-border remote work creates both labor-market opportunity and governance complexity, and that cybersecurity roles amplify this complexity due to privileged access, continuous monitoring needs, and heightened security and privacy constraints [10], [13]–[17], [20]. At the same time, empirical cyberlaw work often lacks direct legal outcome labels, motivating the use of proxy-based measurement strategies and unsupervised typology discovery to structure analysis [18], [19], [21]. The most defensible framing, therefore, is to treat discovered clusters as empirical configurations that may correlate with differing coordination burdens—not as determinations of applicable law, liability, or compliance.

Methodologically, the reviewed work supports interpretable clustering on mixed-type features, careful validation, and explicit robustness and uncertainty reporting [18], [23]. It also implies practical design choices for typology studies: define jurisdictional complexity as multi-dimensional; include remote intensity, organizational scale, and cross-border structure as observable proxies; and pair quantitative typologies with policy/legal interpretation and sensitivity analyses to avoid over-claiming [10], [18], [21], [23]. These principles directly motivate an interpretable mixed-data clustering pipeline with explicit micro-cluster handling, suitable as an empirical entry point for cyberlaw discussion when richer legal variables are unavailable.

Method

Data source, scope, and cleaning

The study uses a single structured salary dataset containing annual records of cybersecurity-related employment arrangements and compensation. Each record includes year (work_year), experience level, employment type, job title, salary in USD (salary_in_usd), employee residence, company location, remote-work intensity (remote_ratio), and company size. The analytical objective is not to predict outcomes but to discover recurring cross-border remote-work patterns that plausibly entail different levels of jurisdictional coordination (e.g., employment-law, payroll/tax, and data-handling governance) using only variables available in the dataset.

We first validate the schema by checking the presence of required columns and enforcing basic type consistency. Salary_in_usd and remote_ratio are converted to numeric types with invalid values coerced to missing. Rows missing either of these fields are removed to prevent distortions in compensation summaries and remote-work categorization. Exact duplicate rows are removed to reduce over-counting of identical employment records and to stabilize distance-based clustering.

Next, we derive a cross-border indicator, cross_border, defined as whether the employee's residence differs from the company's location (employee_residence != company_location). This binary variable is the core operationalization of cross-border status. We also map remote_ratio into a categorical label remote_cat using the dataset's conventional coding: 0→Onsite, 50→Hybrid, and 100→Remote (with any other values retained as "Other" if present). Finally, for distributional robustness and later optional modeling checks, we compute log_salary_usd = log(1 + salary_in_usd), which reduces the influence of extreme salary outliers.

Because the research question concerns jurisdictional complexity in cross-border arrangements, the clustering stage is applied to the cross-border subset only. This two-stage design avoids a dominant “domestic vs cross-border” split that can occur when cross-border cases are rare. The domestic subset is retained for context and descriptive comparison, but the typology is learned from records where cross-border conditions are explicitly present.

Feature engineering and technical representation

Clustering is performed on a mixed-type feature set designed to capture factors plausibly linked to jurisdictional coordination: remote intensity (remote_ratio), employment type (employment_type), seniority proxy (experience_level), firm scale (company_size), and optionally role semantics and jurisdiction categories. To incorporate job-title information without introducing high-dimensional sparse text, we transform job_title into a coarse role_group using transparent keyword rules (e.g., Manager/Lead, Analyst/SOC, Architect, IR/Forensics). This rule-based grouping improves interpretability while keeping the feature space small and stable.

To make the jurisdiction dimension explicit, we include employee_residence and company_location as categorical features rather than relying only on the binary cross_border flag. This preserves which jurisdictions participate in cross-border relationships and enables the clusters to reflect whether the cross-border set is concentrated in a few corridors or spread across many pairs. During reporting, we also derive a “pair” label in the form residence→company_location to quantify corridor diversity within each cluster.

Because the feature set mixes numeric and categorical variables, we use Gower distance to compute pairwise dissimilarities. For numeric variables (here, remote_ratio), distances are computed as normalized absolute differences scaled by the variable’s range. For categorical variables (e.g., employment type, experience level, company size, role group, and jurisdictions), distances are 0 when values match and 1 when they differ. The overall distance between two records is the average across the selected features, producing a bounded distance matrix suitable for downstream clustering.

This implementation produces a full precomputed distance matrix D with shape N×N for the cross-border subset (N = number of cross-border records). The matrix is stored as a checkpoint (.npy) to ensure reproducibility and to support alternative clustering methods (hierarchical or density-based) without recomputing distances. The use of a precomputed distance matrix also allows consistent internal validation using silhouette score computed directly from D.

Unsupervised clustering with minimum cluster size enforcement

We apply agglomerative hierarchical clustering using average linkage on the precomputed Gower distance matrix. Average linkage merges clusters based on mean inter-cluster distance and tends to yield balanced, interpretable groupings for mixed data. The hierarchical procedure produces a linkage matrix from which flat clusterings can be extracted for different candidate numbers of clusters (K). For transparency, we optionally generate a truncated dendrogram, which visually summarizes the merge structure without overplotting large numbers of leaves.

Model selection is performed by evaluating K over a bounded range (default K=2…12), automatically clipped to N-1 for feasibility. For each K, we obtain initial cluster assignments using a max-cluster cut. We then compute the silhouette score using the precomputed distance matrix, which measures how well each record fits within its cluster compared with its nearest alternative cluster. When outliers are enabled, silhouette is computed on the non-outlier subset only, ensuring the metric reflects cluster cohesion rather than being dominated by noise points.

To prevent unstable "micro-clusters" from being interpreted as substantive typologies, we enforce a minimum cluster size threshold (default MIN_CLUSTER_SIZE = 5). Two policy options are supported. In the merge policy, any cluster with size < 5 is reassigned to the nearest large cluster, where "nearest" is defined as the large cluster with the smallest mean Gower distance to the micro-cluster's members. In the outlier policy, all members of micro-clusters are labeled as noise (-1) and excluded from cluster-based typology claims. This enforcement step is applied after the initial hierarchical cut and before final model selection.

The final K is selected by maximizing the silhouette score, with tie-breaking rules that prefer fewer clusters and lower noise fractions when applicable. After selecting K, we relabel clusters into compact indices (0…K−1), preserving -1 for outliers if that policy is used. The labeled cross-border dataset is saved as a checkpoint and also merged back into the full dataset under cb_cluster, where domestic records receive cb_cluster = -1 by construction.

Cluster characterization, compensation analysis, and reproducibility

Cluster interpretation is grounded in descriptive profiling. For each cluster we compute: size (n), fully-remote prevalence (remote100_rate), contractor/freelance prevalence (ctfl_rate, based on employment type in {CT, FL}), company-size composition (shares of S/M/L), and compensation summaries (median, mean, and interquartile range of salary_in_usd). To capture jurisdictional structure, we also report the number of unique residences, unique company locations, and unique residence→company pairs (uniq_pairs). These statistics support a data-driven typology describing how cross-border work is organized.

To connect clusters to the cyberlaw motivation without overclaiming legal conclusions, we use jurisdictional diversity and remote intensity as observable proxies for coordination burden. For example, clusters with high fully-remote rates, high large-firm share, and high corridor diversity (many unique residence→company pairs) are interpreted as settings more likely to require multi-jurisdiction policy alignment and governance processes. This interpretation remains strictly grounded in dataset features; it does not infer compliance status or legal risk for any specific jurisdiction pair.

Where sample sizes permit, we examine whether compensation differs across discovered clusters. Because salary distributions are typically skewed, the workflow supports nonparametric testing (Kruskal–Wallis across clusters, and pairwise Mann–Whitney U tests with Holm–Bonferroni correction) on salary_in_usd. These tests are treated as exploratory, especially when clusters are small, and are complemented by visual summaries such as log-scaled boxplots. The analysis is designed to report effect patterns (e.g., medians and

IQRs) even when formal significance is limited by sample size.

Result and Discussion

Interpretation of Findings

After cleaning, the dataset contained 1,162 unique observations, with 85 exact duplicates removed. Because cross-border records are a minority of the full sample, clustering the full dataset tends to produce a dominant split between domestic and cross-border work. To ensure the typology specifically reflects cross-border arrangements, we apply the unsupervised clustering pipeline to the cross-border subset only (defined by employee_residence ≠ company_location) and then attach the resulting cross-border cluster labels back to the full dataset for contextual reporting.

Hierarchical agglomerative clustering was performed on a precomputed Gower distance matrix constructed from a mixed feature set capturing remote intensity (remote_ratio), job arrangement and seniority (employment_type, experience_level), firm scale (company_size), and—when enabled—role semantics (role_group) and jurisdiction categories (employee_residence, company_location). Model selection evaluated candidate solutions over a bounded range of cluster counts and used silhouette score computed on the precomputed distance matrix. To avoid over-interpreting rare patterns as stable typologies, we enforced a minimum cluster size threshold (n≥5), merging micro-clusters into the nearest larger cluster by mean inter-cluster Gower distance.

The final minimum-size–compliant solution produced two cross-border clusters. Cluster 0 is the dominant configuration with n=52 records, while Cluster 1 contains n=6 records. This indicates that cross-border employment in the dataset is primarily organized around one prevalent pattern, with a smaller secondary pattern distinguished by different remote intensity and organizational scale.

Cluster 0 (n=52) is characterized by a strong “remote-first enterprise” profile. Fully remote work is common, with remote100_rate = 0.8269, and the cluster is overwhelmingly concentrated in large organizations (share_company_L = 0.8846). Contractor/freelance arrangements are rare (ctfl_rate = 0.0192), indicating that this configuration largely reflects employee-type cross-border relationships rather than independent contracting. Compensation within Cluster 0 is heterogeneous: the median salary is $91,886, the mean is $110,235.60, and the salary interquartile range spans $55,666.00 to $150,916.75, consistent with substantial dispersion across roles and seniority.

Cluster 1 (n=6) reflects a smaller-company cross-border configuration with mixed remote arrangements. Its fully remote prevalence is markedly lower (remote100_rate = 0.5000) than Cluster 0. In contrast to the enterprise dominance of Cluster 0, Cluster 1 has no large-firm representation (share_company_L = 0.0000) and is concentrated in mid-sized and small firms (share_company_M = 0.6667, share_company_S = 0.3333). The cluster contains no contractor/freelance cases (ctfl_rate = 0.0000). Compensation in Cluster 1 is comparatively compact, with a median salary of $85,483, a mean of $94,143.00, and a narrow interquartile range of $80,409.25 to $96,491.50.

The clusters also differ in the breadth of jurisdictional corridors represented. Cluster 0 spans 24 distinct employee residences and 17 distinct company locations, resulting in 38 unique residence→company jurisdiction pairs (uniq_pairs = 38). Cluster 1 contains 6 unique residence→company pairs across 6 residences and 6 company locations, indicating that each case corresponds to a distinct corridor but within a small, firm-scale–constrained configuration. Together, these statistics show that the dominant cross-border mode in the dataset is not tied to a single corridor; rather, it recurs across many distinct jurisdiction pairs.

The visualization outputs support the profile-driven interpretation. Figure 1 (cross-border hierarchical dendrogram, truncated) summarizes the agglomeration structure produced by average-linkage clustering on the Gower distance matrix. The dendrogram indicates that many cross-border records form locally cohesive groups that merge at moderate dissimilarity levels (roughly in the 0.4–0.6 range), while the final merges occur closer to the top of the scale (approximately 0.7–0.8). This pattern is consistent with a dataset that contains several small, tight substructures but only a limited number of stable, higher-level partitions once minimum cluster size constraints are enforced.

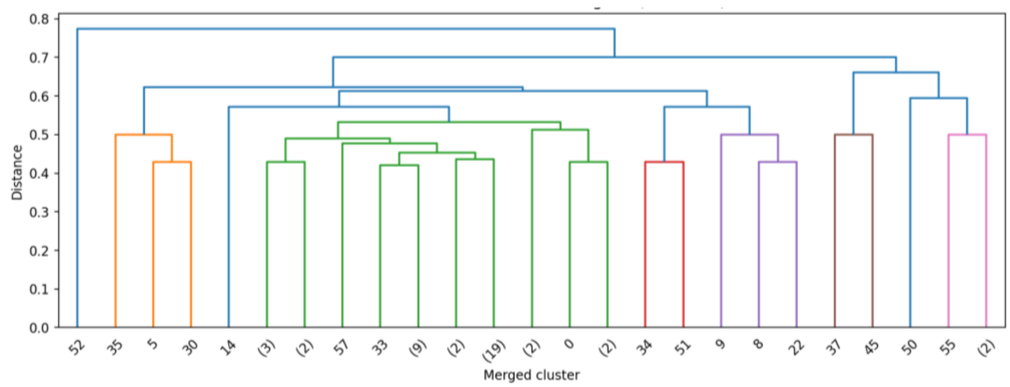

Figure 1 Cross-border hierarchical dendogram

Figure 2 (salary by cluster, log scale) complements the tabular profiles by making compensation dispersion visually explicit. Cluster 0 shows a broad interquartile range and long whiskers, indicating substantial heterogeneity and the presence of lower- and higher-end salaries within the dominant cross-border configuration. Cluster 1 shows a compact box with short whiskers, consistent with the tight IQR reported in the cluster profile table.

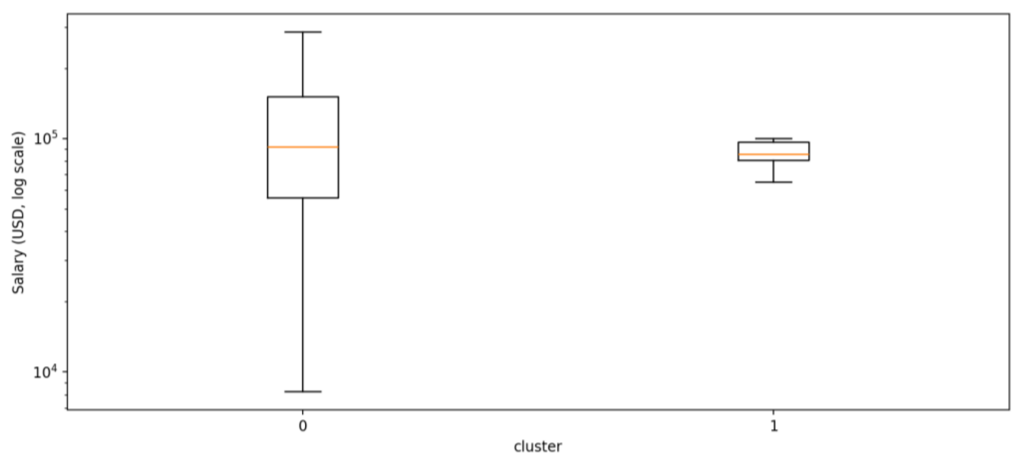

Figure 1 Confusion Matrix

In combination, Figures 1–2 support the interpretation that the principal separation is not simply "high vs low salary" but rather a structural distinction (enterprise remote-first versus mid/small mixed-remote) that is accompanied by different salary dispersion patterns. Cluster size plots indicate that Cluster 0 dominates the cross-border subset, while Cluster 1 forms a small but minimum-size–compliant secondary group. The cluster profile heatmap highlights separation driven primarily by organizational scale (large versus mid/small) and remote intensity (fully remote dominant versus mixed). Salary boxplots (log scale) show wider dispersion for Cluster 0 and a tighter distribution for Cluster 1, consistent with the reported interquartile ranges.

Discussion

The typology indicates that, in this dataset, cross-border work arrangements are structured mainly by two observable factors: remote intensity and organizational scale, rather than by contractor versus employee status. The dominant cluster combines high fully remote prevalence with strong large-firm concentration, suggesting that cross-border remote work is most frequently represented as an enterprise-supported configuration. From a cyberlaw perspective, large organizations operating across many residence→company corridors plausibly require broader coordination of jurisdiction-sensitive processes, such as cross-border policy alignment, access governance, and standard operating procedures for distributed work. While the dataset cannot measure compliance directly, the observed configuration provides an empirical foundation for discussing where coordination burdens are likely to be concentrated.

Cluster 0's corridor diversity (uniq_pairs = 38) is a critical empirical signal for the paper's "jurisdictional complexity" framing. The combination of high fully remote prevalence and many distinct corridors suggests a setting where cross-jurisdiction interactions may be routine and standardized, rather than occasional exceptions. This supports interpreting Cluster 0 as a "high breadth" cross-border mode: the same organizational pattern repeats across multiple residence–company pairs, which is consistent with the need for scalable governance and repeatable processes. The wide salary dispersion further indicates that this mode spans multiple job functions and seniority levels, implying that cross-border remote work is not limited to a narrow role category.

Cluster 1, in contrast, represents a smaller-company cross-border configuration with mixed remote arrangements and a tighter compensation band. A plausible explanation is that mid/small firms may support cross-border work less uniformly, resulting in fewer instances and a more homogeneous compensation range in the observed sample. Because Cluster 1 contains only six observations, its interpretation should be treated as exploratory. However, its consistent separation from Cluster 0 by company size and remote intensity suggests that it captures a distinct configuration in the available feature space rather than being a random split.

Methodologically, enforcing a minimum cluster size is essential for preventing over-interpretation of rare patterns in an imbalanced cross-border subset. Without minimum-size enforcement, hierarchical clustering can yield singleton or very small clusters that are unstable and not suitable for typology claims or comparative analysis. By merging micro-clusters into the nearest larger cluster based on mean Gower distance, the analysis preserves coverage while ensuring each reported cluster reflects a pattern supported by multiple observations. This design choice is particularly important when translating unsupervised groupings into policy-relevant categories.

Limitations

Several limitations bound the conclusions. The cross-border indicator is derived from a mismatch between residence and company location; it does not capture legal place of work, physical presence, multi-country residence histories, or other legal determinants. The dataset also lacks direct variables on regulatory regimes, tax status, work authorization, or data localization requirements, meaning the analysis cannot claim legal risk or noncompliance. In addition, the role grouping is a coarse mapping of job titles and may obscure finer occupational differences that could further structure cross-border work.

The results nonetheless provide a defensible empirical entry point for cyberlaw discussion: the dominant cross-border arrangement in the dataset is enterprise-supported, remote-first, and corridor-diverse, while a smaller configuration involves mid/small firms with mixed remote intensity and tighter compensation. Future work can strengthen robustness through sensitivity analyses that vary minimum cluster size thresholds, toggle inclusion of jurisdiction categories and role grouping, and compare hierarchical clustering with a density-based alternative (e.g., DBSCAN) that naturally labels low-density cases as outliers. Reporting the stability of the enterprise-dominant pattern under these variations would reinforce the credibility of the typology.

Finally, the typology can be used to motivate downstream, dataset-consistent analyses aligned with the cyberlaw theme. Examples include measuring corridor concentration (whether a few residence→company pairs dominate), examining temporal shifts by work_year in the composition of the dominant cluster, and comparing compensation distributions across typology groups as descriptive indicators of how cross-border configurations differ in labor-market outcomes. By maintaining a strict boundary between observable configuration features and legal inferences, the study demonstrates how machine learning can support exploratory mapping of cross-border remote work arrangements in cyberlaw contexts.

Conclusion

This study proposed an unsupervised, dataset-only approach to characterize cross-border remote work arrangements through a jurisdictional-complexity lens. Using a mixed-type representation and Gower-distance hierarchical clustering applied specifically to cross-border cases, we derived a minimum-size–compliant typology that avoids over-interpreting micro-clusters and outliers. The resulting structure indicates that cross-border work in the dataset is dominated by an enterprise-supported, remote-first configuration, while a smaller secondary configuration reflects mid/small-firm arrangements with mixed remote intensity and a tighter compensation band. Substantively, the findings suggest that observable configuration factors—especially remote intensity and organizational scale—are stronger separators of cross-border work patterns than contractor status in this dataset. The dominant cluster spans many residence→company corridors, supporting its use as an empirical proxy for broader coordination burden, while remaining agnostic about actual legal compliance. Future work should report sensitivity analyses (e.g., alternative minimum cluster thresholds, feature toggles for jurisdictions and role grouping, and comparisons with density-based clustering) and incorporate additional legal and governance variables where available to connect typology membership to concrete cyberlaw outcomes.

Declarations

Author Contributions

Conceptualization: A.S.B.; Methodology: M.A.A.; Software: A.S.B.; Validation: M.A.A.; Formal Analysis: M.A.A.; Investigation: A.S.B.; Resources: M.A.A.; Data Curation: M.A.A.; Writing Original Draft Preparation: A.S.B.; Writing Review and Editing: M.A.A.; Visualization: A.S.B.; All authors have read and agreed to the published version of the manuscript.

Data Availability Statement

The data presented in this study are available on request from the corresponding author.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] A. H. Mohammed, M. Rakibuzzaman, and M. A. Alam, “The Future of Work in Financial and Cybersecurity Domains: Analyst Perspectives,” *Journal of Computer Science and Technology Studies*, 2025, doi: 10.32996/jcsts.2025.7.7.79.

[2] Z. Sava, “The Global Shift Towards Remote Work and Its Legal Implications,” *Interdisciplinary Journal of Research and Development*, 2025, doi: 10.56345/ijrdv12n1s137.

[3] S. Reddy, “Decoupling Cloud Security (DCS): A Framework for Data Sovereignty and Cross-Border Cloud Compliance,” *Journal of Information Systems Engineering & Management*, 2025, doi: 10.52783/jisem.v10i4.9317.

[4] Y. Wu, “Balancing Data Protection and Data Utilization: Global Perspectives and Trends,” *Lecture Notes in Education Psychology and Public Media*, 2024, doi: 10.54254/2753-7048/44/20230095.

[5] J. Hwang, “Digital Sovereignty in an Era of Cyber Threats and Global Connectivity,” *International Journal of Multidisciplinary Research Updates*, 2025, doi: 10.53430/ijmru.2025.9.2.0023.

[6] P. S. Tandberg, “The Conflict Rules of European Social Security Coordination in a Digital Labour Market – Bridging *lex loci laboris* and Remote Work by Disregarding Marginal Cross-Border Activities,” *European Journal of Social Security*, 2025, doi: 10.1177/13882627251319638.

[7] M. I. Khan, “Legal Documentation and Case Management: A Systematic Review of Digitization Trends and Cybersecurity Challenges in Legal Support Roles,” *Rast*, 2023, doi: 10.63125/21hf4w52.

[8] V. Mone, R. Tilwani, C. Sivakumar, and S. Fayzullaeva, “Evaluating the Prospects of a UN-Backed Global Data Protection Authority: A Third World Perspective,” *International Organizations Law Review*, 2025, doi: 10.1163/15723747-22010002.

[9] T. D. Allen, T. D. Golden, and K. M. Shockley, “How Effective Is Telecommuting? Assessing the Status of Our Scientific Findings,” *Psychological Science in the Public Interest*, 2015, doi: 10.1177/1529100615593273.

[10] Y. Elsamani and Y. Kajikawa, “How Teleworking Adoption Is Changing the Labor Market and Workforce Dynamics?,” *PLOS ONE*, 2024, doi: 10.1371/journal.pone.0299051.

[11] M. Kwon and M. Kim-Goh, “The Impacts of Telework Options on Worker Outcomes in Local Government: Social Exchange and Social Exclusion Perspectives,” *Review of Public Personnel Administration*, 2022, doi: 10.1177/0734371x221121051.

[12] J. L. O. Beckel and G. G. Fisher, “Telework and Worker Health and Well-Being: A Review and Recommendations for Research and Practice,” *International Journal of Environmental Research and Public Health*, 2022, doi: 10.3390/ijerph19073879.

[13] M. O. Aduwo, A. B. Akonobi, and C. O. Makata, “Cybersecurity and Data Privacy Compliance Framework for Human Resource Analytics and Digital Workforce Information Systems,” *Int J Sci Res Humanities and Social Sciences*, 2024, doi: 10.32628/ijsrssh242658.

[14] A. F. Al-Qahtani and S. Cresci, “The COVID-19 Scamdemic: A Survey of Phishing Attacks and Their Countermeasures During COVID-19,” *IET Information Security*, 2022, doi: 10.1049/ise2.12073.

[15] A. Mihailović, J. C. Smolović, I. Radević, N. Rašović, and N. Martinović, “COVID-19 and Beyond: Employee Perceptions of the Efficiency of Teleworking and Its Cybersecurity Implications,” *Sustainability*, 2021, doi: 10.3390/su13126750.

[16] O. Ejiofor, O. Olusoga, and A. Akinsola, “Zero Trust Architecture: A Paradigm Shift in Network Security,” *Computer Science & IT Research Journal*, 2025, doi: 10.51594/csitrj.v6i3.1871.

[17] N. Parvatha, “Securing Multi-Tenant Cloud Platforms During Global Crises: A Zero Trust Approach,” *International Journal of Science and Research Archive*, 2020, doi: 10.30574/ijsra.2020.1.1.0017.

[18] E. C. Matthay *et al.*, “The Revolution Will Be Hard to Evaluate: How Co-Occurring Policy Changes Affect Research on the Health Effects of Social Policies,” *Epidemiologic Reviews*, 2021, doi: 10.1093/epirev/mxab009.

[19] J. L. Jahn *et al.*, “Legislating Inequity: Structural Racism in Groups of State Laws and Associations With Premature Mortality Rates,” *Health Affairs*, 2023, doi: 10.1377/hlthaff.2023.00471.

[20] M. Clarke and K. Martin, “Managing Cybersecurity Risk in Healthcare Settings,” *Healthcare Management Forum*, 2023, doi: 10.1177/08404704231195804.

[21] D. Ezzine-de-Blas, S. Wunder, M. Ruíz-Pérez, and R. del Pilar Moreno-Sánchez, “Global Patterns in the Implementation of Payments for Environmental Services,” *PLOS ONE*, 2016, doi: 10.1371/journal.pone.0149847.

[22] E. Molin, P. L. Mokhtarian, and M. Kroesen, “Multimodal Travel Groups and Attitudes: A Latent Class Cluster Analysis of Dutch Travelers,” *Transportation Research Part A: Policy and Practice*, 2016, doi: 10.1016/j.tra.2015.11.001.

[23] R. Facendola, G. O. Palmisano, A. D. Boni, C. Acciani, and R. Roma, “Profiling Citizens on Perception of Key Factors of Food Security: An Application of K-Means Cluster Analysis,” *Sustainability*, 2023, doi: 10.3390/su15139915.

[24] M. D. Angelis, L. Mantecchini, and L. Pietrantoni, “A Cluster Analysis of University Commuters: Attitudes, Personal Norms and Constraints, and Travel Satisfaction,” *International Journal of Environmental Research and Public Health*, 2021, doi: 10.3390/ijerph18094592.