computers

MDPI

Article

An Intelligent Data-Driven Framework for Web Accessibility Diagnosis and Remediation

Patricia Acosta-Vargas [1,*], Daniel Diaz-Bedoya [1], Luis Salvador-Ullauri [2], Sylvia Novillo-Villegas [1], Belén Salvador-Acosta [3] and Mario González-Rodríguez [1]

[1] Intelligent and Interactive Systems Laboratory, Universidad de Las Américas, Quito 170503, Ecuador; daniel.diaz.bedoya@udla.edu.ec (D.D.-B.); sylvia.novillo@udla.edu.ec (S.N.-V.); mario.gonzalez.rodriguez@udla.edu.ec (M.G.-R.)
[2] Department of Software and Computing Systems, University of Alicante, 03690 Alicante, Spain; lasu1@alu.ua.es
[3] Facultad de Medicina, Universidad de las Américas, Quito 170122, Ecuador; maria.salvador.acosta@udla.edu.ec
* Correspondence: patricia.acosta@udla.edu.ec

Abstract

Web accessibility assessment remains challenging because of the scale and heterogeneity of modern institutional websites and the difficulty of translating automatically detected violations into actionable remediation priorities. This paper introduces the Data-Driven Web Accessibility Remediation Framework (DD-WARF), a computational framework that integrates automated WCAG-based accessibility evaluation with statistical analysis, density-based clustering, similarity network analysis, and quantitative remediation modeling. DD-WARF combines DBSCAN-based anomaly detection to identify atypical accessibility profiles, cosine-similarity networks to characterize recurring patterns across institutions, and a Weighted Migration Effort Index (WMEI) to estimate relative remediation effort and support intervention prioritization. The framework was empirically evaluated using accessibility data from 320 Latin American university websites assessed with Google Lighthouse and WAVE under WCAG 2.2. The analysis revealed no statistically significant association between institutional ranking and the number of automatically detected accessibility errors ($\rho = -0.0961$, $p = 0.086$). Three WCAG success criteria, 1.4.3 (Contrast (Minimum)), 1.1.1 (Non-text Content), and 2.4.4 (Link Purpose (In Context)), emerged as the dominant accessibility barriers, affecting nearly 90% of the evaluated institutions. DBSCAN further identified nine institutions with atypical accessibility profiles, while similarity network analysis revealed recurring accessibility patterns shared across institutions. By applying a simulated remediation scenario targeting these dominant criteria, the DD-WARF framework projects a theoretical accessibility improvement ranging from 37.4% to 70.6% across institutional profiles. These findings show the potential of DD-WARF as a reproducible, data-driven decision-support framework for large-scale accessibility diagnosis, benchmarking, remediation prioritization, and resource allocation.

Keywords: web accessibility; WCAG; data mining; DBSCAN; similarity network analysis; accessibility remediation; decision support; higher education; digital inclusion

Check for updates

Academic Editor: Makuochi Samuel Nkwo

Received: 24 August 2026
Revised: 1 October 2026
Accepted: 5 October 2026
Published: 8 October 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Web accessibility remains a global challenge, with a high percentage of websites failing to meet WCAG (Web Content Accessibility Guidelines) [1]. The digital divide in higher

education in Latin America is a multifaceted problem, and web accessibility remains a fundamental issue that is inadequately addressed [2].

Analyses of Latin American university websites have revealed significant accessibility issues, such as a lack of alternative text, contrast errors, and navigation difficulties. For example, a study of the top 20 universities identified 434 contrast errors, 210 missing alternative texts, and 262 navigation problems, even at institutions with high global rankings such as USP and UNAM. Similarly, an evaluation of 348 university websites found frequent violations of the Web Content Accessibility Guidelines (WCAG) 2.0, 2.1, particularly regarding the inclusion of alternative text for images [3].

Inequalities in digital inclusion: High scores in technological innovation, such as the adoption of artificial intelligence, do not necessarily correlate with better web accessibility. This underscores the need for integrated strategies that prioritize inclusion along with technological progress [4].

Broader implications of the digital divide: Socioeconomic and regional inequalities: The digital divide is exacerbated by socioeconomic disparities, with rural and low-income students facing greater difficulties in accessing digital resources. For example, during the COVID-19 pandemic, students in rural areas of countries such as Peru and Guatemala experienced problems with internet connectivity and access to digital devices, further widening the gap in educational opportunities [5].

Impact on student outcomes: The digital divide significantly affects learning and retention in higher education [6]. Structural limitations, such as poor infrastructure and inadequate digital training programs, contribute to high dropout rates, particularly among socioeconomically disadvantaged students. Recommendations for Addressing the Gap: Adoption of Accessibility Standards: Universities should align with WCAG 2.2 and integrate accessibility into the early stages of website development to foster inclusion and improve user experience [7].

Policy and Training Initiatives: Strengthening public policies and implementing digital literacy programs are essential for closing this gap. Contextualized research and digital skills training can help mitigate inequalities. Systemic Interventions: Addressing the digital divide requires systemic approaches, including investments in infrastructure, targeted support for rural and low-income students, and promoting universal design principles [8].

While Latin American universities are making progress in digital visibility, significant gaps persist in web accessibility and digital inclusion [9]. Addressing these issues is fundamental to achieving equitable access to higher education and aligning with the Sustainable Development Goals (SDGs) related to quality education and reducing inequalities [10]. To promote transparency and reproducibility, we publicly released the resulting accessibility dataset through Mendeley Data [11].

To address the limitations of current large-scale accessibility assessment approaches, this study proposes the Data-Driven Web Accessibility Remediation Framework (DD-WARF), a scalable computational framework that integrates automated WCAG-based evaluation with statistical analysis, density-based clustering, and similarity network analysis. Unlike conventional accessibility audits that primarily report compliance violations, the proposed framework provides a structured methodology for identifying accessibility patterns, classifying institutional websites, and prioritizing remediation strategies.

In this study, "data-driven" means that remediation priorities are derived from measured patterns of automatically detected accessibility errors across websites, rather than from an isolated list of violations for each website. Conventional automated testing tools identify machine-detectable issues and report their locations or counts. DD-WARF uses those outputs as input for cross-website statistical analysis, outlier detection, similarity network analysis, website-profile classification, and quantitative estimation of relative

remediation effort. These additional steps support comparison and prioritization; they do not replace manual inspection or evaluation with users with disabilities. A key contribution of this work is the Weighted Migration Effort Index (WMEI), a quantitative metric that estimates the technical effort required to remediate accessibility barriers based on the complexity of different WCAG violations. By combining DBSCAN clustering with graph-based similarity analysis, the framework identifies structural accessibility outliers, characterizes common accessibility profiles, and supports evidence-based decision-making for developers and institutional stakeholders.

The proposed methodology was empirically evaluated through a large-scale analysis of 320 university websites across Latin America, demonstrating its applicability to real-world web ecosystems. Although evaluated in the higher education domain, the framework is designed as a reproducible and generalizable approach that can be adapted to other large-scale web environments requiring automated accessibility assessment, benchmarking, and remediation planning. The main contributions of this study are fourfold: (1) the development of DD-WARF, a reproducible computational framework integrating automated WCAG evaluation, statistical analysis, density-based clustering, and similarity network analysis; (2) the identification of structural accessibility archetypes through DBSCAN and graph-based community detection; (3) the introduction of the Weighted Migration Effort Index (WMEI), which transforms accessibility profiles into quantitative remediation-effort estimates; and (4) the empirical evaluation of the framework using 320 Latin American university websites, providing a large-scale assessment of its applicability to real-world web ecosystems.

2. Related Work

2.1. Automated Web Accessibility Evaluation

Definition and Purpose: Automated web accessibility evaluation involves using software tools to assess whether web content complies with accessibility guidelines, such as the Web Content Accessibility Guidelines (WCAG) [12,13]. Tools and Methods: These tools (often called Web Accessibility Evaluation Tools, WAET) scan web pages for errors and barriers, providing fast and responsive feedback [14,15]. They are widely used to reduce the burden of manual inspection, which is time-consuming and requires expertise [16]. Limitations: Automated tools typically detect only a subset of accessibility issues, mainly those identifiable through code analysis. They struggle with rules requiring semantic understanding or user experience evaluation, often returning binary results (error/no error) and missing nuanced problems [17]. Complementary Role: Experts recommend combining automated tools with manual and user-based evaluations for a thorough assessment.

2.2. Data-Driven Accessibility Analysis

Approach: Data-driven accessibility analysis leverages large-scale data collection and analytics to assess accessibility trends and identify common barriers [18]. Examples: Studies have analyzed millions of web pages to quantify error prevalence and characterize the overall state of web accessibility, revealing frequent issues like poor text contrast and missing alternative text. Broader Applications: Data-driven methods can also incorporate geographic and demographic data to assess accessibility in specific contexts, such as public transportation systems [18].

2.3. Machine Learning and AI for Accessibility

Emerging Role: Machine learning (ML) and artificial intelligence (AI) are increasingly used to address limitations of traditional automated tools, especially for tasks requiring semantic understanding or interpretation of complex content [19]. Applications: ML techniques can

enhance navigation for users with disabilities, interpret media content, and power assistive technologies like OCR and text-to-speech devices [20]. Recent Advances: Large language models (LLMs) show promise in automating tasks previously thought to require human expertise, potentially expanding the scope of automated accessibility evaluation.

2.4. Web Accessibility and Information System Quality

Web accessibility fits within a broader account of information system (IS) quality. A proposed 20-year update of the IS success model introduces social quality and a hierarchy of quality dimensions, offering a conceptual basis for examining how an information system serves different users and social needs [21].

From this perspective, accessibility has implications for several related dimensions: the ability to perceive and operate an interface concerns system quality; the availability and clarity of content concern information quality; and equitable access to digital services is relevant to social quality. These relationships provide a rationale for treating accessibility as more than a count of technical violations. They do not, however, imply that automated WCAG findings alone measure overall IS success or establish a validated hierarchy of these dimensions.

AI-assisted remediation also raises questions about whether stakeholders will accept and use the resulting recommendations. The revised artificial intelligence device use acceptance (RAIDUA) model highlights the role of privacy concerns in accepting AI-enabled devices [22]. Although its empirical context differs from web accessibility remediation, it provides a relevant perspective for future evaluation of the AI-assisted recommendation component proposed in DD-WARF. The present study focuses on accessibility diagnosis and model-based remediation planning; it does not test RAIDUA constructs or measure stakeholder acceptance.

2.5. Research Gap and Comparison with Existing Approaches

Overall, automated accessibility evaluation tools provide scalable mechanisms for identifying machine-detectable WCAG violations but remain limited in detecting semantic, contextual, and experiential barriers. Data-driven and AI-based approaches extend these capabilities by enabling large-scale pattern discovery and contextual analysis; however, they do not eliminate the need for expert and user-based evaluation. These limitations highlight the need for integrated computational frameworks that combine automated diagnosis, pattern discovery, prioritization, and remediation support while maintaining human-centered validation. Table 1 summarizes existing approaches to web accessibility assessment and the remaining research gaps.

Table 1. Comparison of Existing Approaches for Web Accessibility Assessment and Remaining Research Gaps.

Aspect	Automated Tools	Data-Driven Analysis	ML/AI Approaches	Remaining Research Gaps
Scope of Issues Detected	Primarily machine-detectable code and WCAG violations	Identifies large-scale trends, patterns, and anomalies	Expanding toward semantic and contextual accessibility analysis	Limited assessment of semantic, contextual, and experiential barriers
Speed and Scalability	High	High for large datasets	Increasing with advances in automated AI-based analysis	Need for scalable integration of complementary assessment methods
Need for Human Evaluation	High for contextual and usability-related issues	Moderate for interpretation and validation	Still required to validate AI-generated assessments and recommendations	Need for user-centered evaluation involving people with disabilities
Current Limitations	Cannot detect all complex or context-dependent accessibility barriers	May identify patterns without explaining their underlying causes	AI-based accessibility assessment and remediation remain emerging areas	Lack of comprehensive frameworks integrating diagnosis, pattern discovery, prioritization, and remediation support

3. Methodology

3.1. Automated Accessibility Evaluation Protocol

To ensure a systematic and reproducible assessment of web accessibility, we established a standardized automated evaluation protocol for the institutional websites included in the study. The complete data collection process was conducted between March and May 2026 and covered 320 university websites across Latin America. For each institution, the official institutional homepage was evaluated using a consistent procedure to minimize variability associated with page selection and evaluation conditions.

To enable a scalable, large-scale accessibility diagnosis, the homepage was designated as the representative proxy because it acts as the primary digital gateway and effectively reflects the structural baseline enforced by centralized Content Management Systems (CMS) and unified design frameworks.

The accessibility assessment employed two complementary automated tools: Google Lighthouse and the WAVE Web Accessibility Evaluation Tool. Google Lighthouse was executed through Google Chrome on Windows in desktop mode, using the stable browser environment available during the evaluation period. The documented final evaluation environment corresponded to Google Chrome 148.0.7778.96/97. WAVE assessments were performed using the WAVE Evaluation Tool browser extension (v3.3.1.0, WebAIM) in Google Chrome on Windows. WAVE analyzes the rendered web page directly within the browser and identifies machine-detectable accessibility barriers associated with WCAG success criteria. For each university, the same institutional homepage URL was evaluated with both tools to maintain consistency and comparability throughout the assessment process. Lighthouse provided a complementary automated accessibility audit, whereas WAVE was used to obtain criterion-level accessibility findings that were subsequently mapped to the corresponding WCAG success criteria.

The assessment focused on machine-detectable accessibility barriers associated with the Web Content Accessibility Guidelines (WCAG 2.2) and conformance levels A and AA. Automated assessments generated accessibility indicators describing the frequency and type of violations detected on each website. The resulting records were then mapped to their corresponding WCAG success criteria to obtain a common representation suitable for computational analysis.

The initial accessibility feature space included the WCAG criteria identified through the automated assessment procedure. Following an exploratory analysis of their prevalence in the dataset, six recurring criteria were retained for regional accessibility characterization: 1.4.3 Contrast (Minimum), 1.1.1 Non-textual content, 2.4.4 Link purpose (In context), 1.3.1 Information and relationships, 2.4.6 Headings and labels, and 3.3.2 Labels or instructions. Among these, criteria 1.4.3, 1.1.1, and 2.4.4 emerged as the dominant accessibility barriers and were subsequently used to construct the three-dimensional accessibility representation employed by DD-WARF for profile characterization and remediation analysis.

We integrated the results from Lighthouse and WAVE into the WCAG success criteria, rather than simply adding the scores generated by each tool. WAVE provided error counts at the criteria level, while Lighthouse provided complementary validation.

The unit of counting was an error occurrence on the evaluated homepage. Multiple occurrences mapped to the same WCAG success criterion on a single homepage were counted separately in that criterion's WAVE error count; for example, three elements with the same detected error contributed a count of three, rather than one binary indication of that error type. Lighthouse findings were used as complementary information and were not added to these counts.

We merged the findings from both tools after mapping them to the WCAG criteria. This harmonization strategy provided a standardized representation of accessibility barriers

across all institutions, while preserving the complementary detection capabilities of both assessment tools.

Before statistical and machine learning analyses, we preprocessed the collected records to ensure data consistency. We removed records corresponding to inaccessible URLs, incomplete evaluations, or missing accessibility measurements. We converted WCAG-related variables into numerical error counts and verified their consistency before analysis. Each institution was represented by an accessibility feature vector, structured and suitable for subsequent statistical analysis, anomaly detection using DBSCAN, similarity network construction, community detection, and remediation effort estimation using WMEI.

This protocol established a consistent computational workflow, from automated accessibility assessment to structured feature extraction, and ensured that the resulting dataset could be systematically analyzed within the DD-WARF framework. The processed dataset is publicly available through Mendeley Data [11] to support reproducibility, benchmarking, and future comparative studies.

3.2. Dataset Collection

The processed dataset used in this study was published in the open-access repository Mendeley Data to ensure scientific transparency, reproducibility, and facilitate future comparative studies [11]. The repository provides permanent access, via DOI, to selected accessibility records from Latin American university websites evaluated using the Lighthouse [23] and WAVE tool [24]. The dataset's public availability enables replication of the analytical framework, comparative studies, and the development of future AI-assisted accessibility remediation approaches.

3.3. Spearman's Rank Correlation Coefficient (ρ)

It is a non-parametric measure of the statistical dependence between the rankings of two variables. Unlike Pearson's correlation, Spearman's coefficient does not require assumptions of normality or linearity because it operates on the ranks of the data points rather than their absolute values [25].

For a sample of size n, the n raw scores X_i, Y_i are converted to ranks $R(X_i), R(Y_i)$. The Spearman correlation coefficient was calculated as follows:

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)},$$

where $d_i = R(X_i) - R(Y_i)$ is the difference between the ranks of the observations. The resulting coefficient ρ ranges from -1 to $+1$:

- $\rho = +1$: Perfect direct monotonic relationship.
- $\rho = 0$: No monotonic correlation between the variables.
- $\rho = -1$: Perfect inverse monotonic relationship.

3.4. Density-Based Spatial Clustering of Applications with Noise (DBSCAN)

It is a density-based clustering algorithm widely used to identify clusters of arbitrary shapes and isolate statistical outliers in multi-dimensional datasets. DBSCAN defines clusters as dense regions of points separated by regions of lower density, labeling points in low-density areas as noise [26,27].

The algorithm is defined by two hyperparameters: the neighborhood radius (ϵ) and the minimum number of points required to form a dense region ($MinPts$). For any point p in a dataset D, its ϵ-neighborhood, $N_\epsilon(p)$, is defined as the set of points within distance ϵ:

$$N_\epsilon(p) = \{q \in D \mid dist(p, q) \leq \epsilon\}.$$

A point p is classified based on the cardinality of its neighborhood:

- Core Point: If $|N_\epsilon(p)| \geq MinPts$, the point is considered to be at the center of a dense region.
- Border Point: If $|N_\epsilon(p)| < MinPts$ but p is density-reachable from a core point.
- Noise Point (Outlier): If p is neither a core nor a border point, it is classified as noise.

By definition, a noise point is not an effective tool for filtering anomalies that deviate from the data's global or local density patterns.

3.5. Cosine Similarity Measure

It is a metric used to determine the degree of similarity between two nonzero vectors by measuring the cosine of the angle between them [28]. Assess the orientation of the vectors rather than their magnitude.

Given two n-dimensional vectors, $\mathbf{A}$ and $\mathbf{B}$, the cosine similarity S_C is calculated as the dot product of the vectors divided by the product of their magnitudes:

$$S_C(\mathbf{A}, \mathbf{B}) = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \sqrt{\sum_{i=1}^{n} B_i^2}}.$$

The resulting value S_C ranges from -1 (exactly opposite) to 1 (exactly the same), with 0 indicating orthogonality (no similarity).

3.6. Similarity Network Analysis (SNA)

It is a methodology derived from graph theory that is used to model complex systems as a set of nodes and their pairwise relationships. It is defined as a graph $G = (V, E)$, where V represents the set of vertices and E represents the set of edges that connect them based on proximity criteria [29].

The structure of the resulting topology is typically defined by the following components:

- Adjacency Logic: A link $e_{i,j}$ exists between nodes v_i and v_j if their pairwise similarity score exceeds a specified threshold τ.
- Edge Weighting: Links can be weighted to represent connection strength, enabling the use of community detection algorithms and structural metrics.
- Topological Analysis: The network allows for the identification of global properties, such as the Small-World effect, and meso-scale structures, such as clusters or modules, which reveal the underlying organization of the data.

We analyzed two networks. In the country-level network, a node represents a country, and an undirected edge represents a retained Spearman correlation between the countries' profiles of mean frequencies across 13 WCAG error types. In the institution-level network, a node represents an evaluated university homepage, and an undirected edge represents a retained cosine similarity between two institutions' error-count vectors across the six selected WCAG criteria. Edge weights correspond to the respective Spearman correlation or cosine-similarity values. The inclusion rules for each network are specified in Sections 4.3 and 4.7.

3.7. Framework Design

The construction of the DD-WARF combined decisions derived from the observed data with design choices required to structure the framework. The methodological choices were aligned with three analytical objectives: comparing automatically detected accessibility errors across institutions, identifying recurring and atypical error profiles, and supporting exploratory remediation prioritization. Homepage sampling provided a common observation unit for comparison and represents a design-based scope restriction rather than a

validated proxy for whole-site accessibility. Six standardized error-count features were retained for DBSCAN and institutional similarity analysis using a prevalence-based rule. The observed prevalence values were empirical, whereas the inclusion threshold of more than 50% and the rule MinPts $= 2D$ were heuristic choices, where $D = 6$ is the number of retained features. We selected $\varepsilon = 3.0$ based on the observed k-distance curve; this is a dataset-specific setting rather than an externally validated optimum.

For remediation modeling, the three-criterion representation was motivated by the observed concentration of errors in WCAG criteria 1.4.3, 1.1.1, and 2.4.4. Restricting the remediation space to these criteria was a design choice intended to support prioritization, not a comprehensive representation of accessibility. Finally, the WMEI weights of 1, 2, and 3 encode assumed differences in relative technical complexity; they were not calibrated against observed engineering hours or costs. These distinctions clarify how empirical observations informed the framework while identifying the assumptions that require further validation.

To complement this structure, the framework introduced a heuristic measure of relative remediation effort, designed to represent differences in the technical complexity associated with correcting different types of accessibility errors and to support prioritization among profiles. Together, these decisions provide a path from observed accessibility findings to institutional characterization and remediation planning.

4. Results

4.1. Academic Ranking vs. Accessibility Compliance

A Spearman's rank correlation analysis was conducted to assess the association between university ranking position and the total number of automatically detected accessibility errors. The analysis revealed a very weak negative correlation ($\rho = -0.0961$, $p = 0.086$, $n = 320$), which was not statistically significant at the $\alpha = 0.05$ level. These findings indicate that there was insufficient evidence of a monotonic association between institutional ranking position and the number of automatically detected accessibility errors in the analyzed sample.

Figure 1 presents a scatter plot comparing the number of errors on a website against its ranking position. No relationship can be perceived, reinforcing the results obtained by the correlation test.

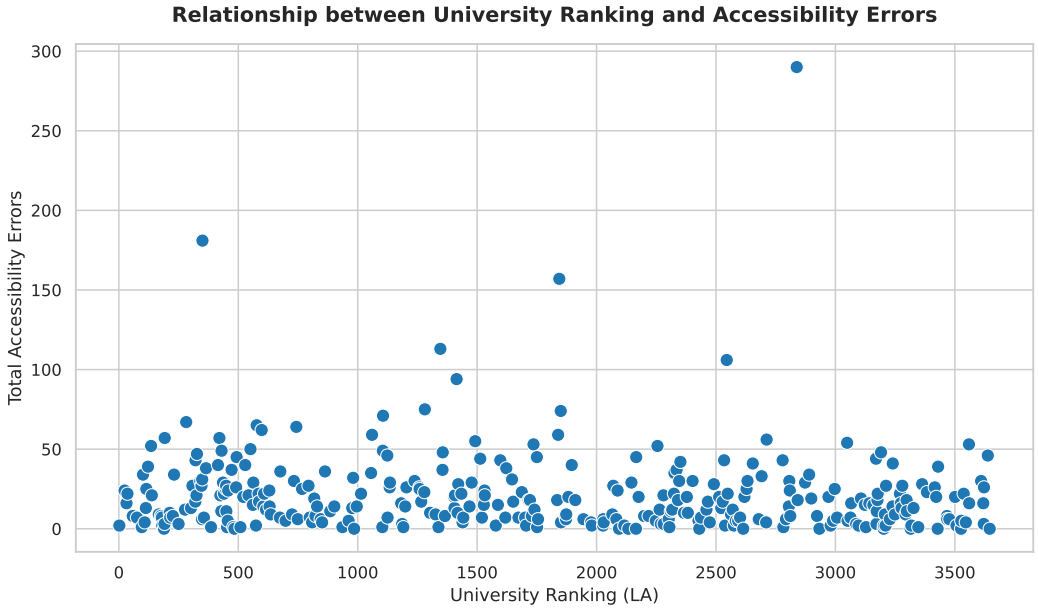


Figure 1. Number of errors compared to the website's ranking.

4.2. Statistical Analysis of Error Distribution by Country

The distribution of the total number of errors by country is shown in Figure 2. Brazil and Mexico have the highest error concentrations in the analysis, but this is because there are more university websites from these two countries.



Figure 2. Number of total errors by country.

Figure 3 shows a box plot of the number of errors by country for the top 10 countries. The color of the box distinguishes countries with more reliable information. The distribution of errors may differ across countries with a small number of analyzed websites. The dotted line indicates the global mean value, calculated by dividing the total number of errors in the region by the number of analyzed websites.

Brazil has most of the analyzed websites below the global mean. However, several extreme observations indicate institutions with substantially higher numbers of accessibility errors. Mexico and Colombia similarly show more than 50% of their evaluated websites below the global mean.

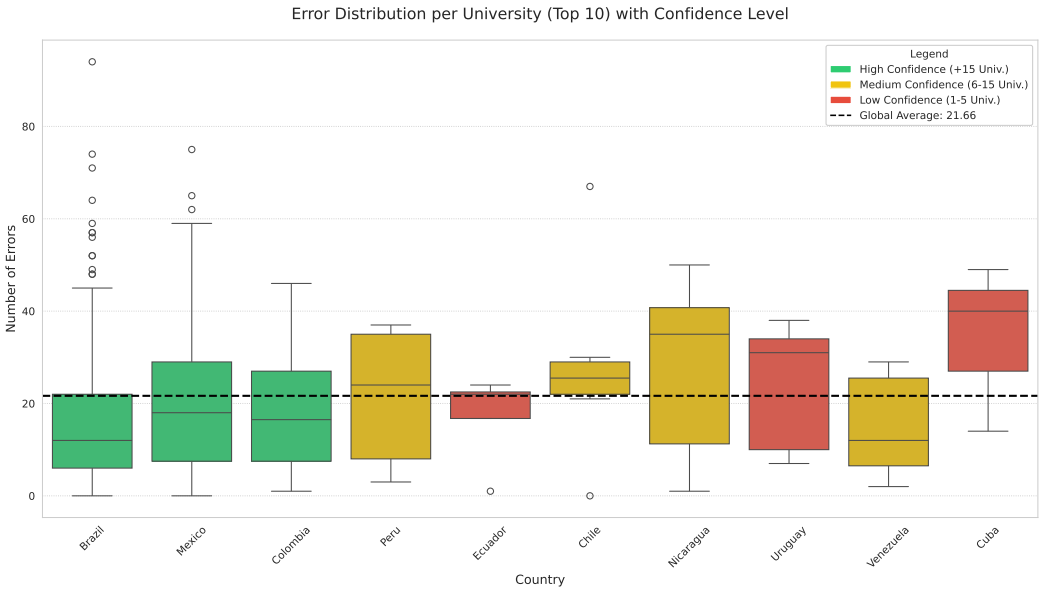


Figure 3. Box plot representation of errors by country.

Chile and Cuba stand out as the two countries with the highest number of websites above the global mean, but both countries need a more extensive analysis because the number of analyzed websites is not enough to have conclusive information. For visual clarity, we excluded five extreme outliers with error counts exceeding 100 from the plot.

Table 2 presents the descriptive statistics of accessibility errors for the ten countries with the highest total error counts. Although the mean number of errors ranges between 15.29 and 36.40, the dispersion metrics reveal substantial skewness across distributions. In Brazil, the median (13.00, IQR = 17.50) is considerably lower than the mean (21.93, Std = 34.42), reflecting the influence of high-error outliers. Also, metrics for countries with very small sample sizes, such as Cuba with $N = 3$, or Ecuador and Uruguay with $N = 5$, must be interpreted with caution. In these cases, the estimates lack statistical reliability and are highly sensitive to individual institutions. A notable example is Ecuador, where a single website with 113 errors substantially inflates both the mean (36.40) and the standard deviation (43.84), despite a median of 22.00 and an IQR of just 2.00. Therefore, expanding the sample size in these underrepresented countries is necessary in future work to achieve representative and robust benchmarks.

Table 2. Descriptive statistics of accessibility errors by country.

Country	Total Websites	Total Errors	Mean Errors	Std Errors	Median Errors	IQR Errors	Min Errors	Max Errors
Brazil	135	2960	21.93	34.42	13.0	17.5	0	290
Mexico	71	1538	21.66	17.74	18.0	21.5	0	75
Colombia	31	671	21.65	20.19	17.0	19.5	1	106
Peru	9	194	21.56	14.05	24.0	27.0	3	37
Ecuador	5	182	36.40	43.84	22.0	2.0	1	113
Chile	6	169	28.17	21.78	25.5	7.0	0	67
Nicaragua	6	167	27.83	20.29	35.0	29.5	1	50
Uruguay	5	120	24.00	14.40	31.0	24.0	7	38
Venezuela	7	107	15.29	11.37	12.0	19.0	2	29
Cuba	3	103	34.33	18.18	40.0	17.5	14	49

To determine whether the observed variations in accessibility errors across countries represent statistically significant differences rather than random sampling noise, an inferential comparative analysis was performed. To ensure statistical robustness, the analysis

focused on countries with a sample size exceeding 10 universities. Consequently, only Brazil ($N = 135$), Mexico ($N = 71$), and Colombia ($N = 31$) met this threshold.

Prior to performing the One-Way ANOVA, several assumptions were tested. First, a Shapiro-Wilk test was applied to each group; the results indicated that the error distributions deviated significantly from normality ($p < 0.001$ for Brazil, Colombia and Mexico). However, given that all groups exceeded 30 observations, the Central Limit Theorem justifies the use of ANOVA due to its robustness against non-normality in large samples.

Second, Levene's test was conducted to assess the homogeneity of variances. The results showed no evidence against the equal variance assumption ($p = 0.636$), fulfilling the prerequisite for parametric testing. The subsequent one-way ANOVA revealed no statistically significant differences in error frequency among the three countries ($p = 0.997$). This lack of statistical disparity indicates that accessibility errors do not differ meaningfully across these nations, suggesting a homogeneous pattern of digital barriers across the analyzed Latin American higher education institutions.

4.3. Cross-Country Error Correlation

To further investigate whether these accessibility barriers are country-specific or represent a systemic regional pattern, a network analysis based on Spearman's Rank Correlation was conducted.

The network analysis was conducted in three stages. First, the dataset was filtered to include only countries with five or more analyzed websites, ensuring that the mean error values per country were statistically robust. Second, the mean frequencies for each of the 13 WCAG error types were calculated for each nation. These values served as the basis for calculating a Spearman Rank Correlation matrix, from which the network was constructed.

We retained an undirected edge when the Spearman correlation between two country profiles was $\rho \geq 0.40$. We excluded negative correlations and correlations below 0.40. In Figure 4, node colors indicate the communities identified by the Louvain algorithm, and edge widths represent the retained correlation coefficients.

To identify clusters of countries with shared accessibility profiles, the Louvain community detection algorithm. This process revealed the emergence of exactly three communities, as illustrated in Figure 4. In this visualization, the edge thickness represents the strength of the correlation between countries, while the node colors distinguish the detected communities.

These results suggest that the observed accessibility profiles are not strictly determined by geographic proximity, as institutions from geographically distant countries may exhibit highly similar patterns of WCAG violations.

An analysis of the top three errors across the detected communities shows that they are the same. As detailed in

Table 3, criteria 1.4.3 (Contrast Minimum), 2.4.4 (Link Purpose), and 1.1.1 (Non-text Content) consistently constitute the top three failure dimensions across all groups in terms of mean incidence.

The descriptive indicators reveal substantial differences in central tendency and dispersion among communities. Community C1 ($N = 213$), which encompasses the largest institutional sample, displays the most stable error profile across all evaluated criteria characterized by lower location parameters and reduced variance. Specifically, for Criterion 1.4.3, the mean reaches 20.20 (Std = 6.19) with a median of 17.71 (IQR = 7.63) and a maximum restricted to 28.67, alongside a similarly restrained dispersion in Criterion 2.4.4 (Std = 3.95) and Criterion 1.1.1 (Std = 4.91). In contrast, Community C0 ($N = 26$) exhibits an intermediate behavior, displaying elevated contrast deficiencies (Criterion 1.4.3: Mean = 33.48, Median = 28.92, reaching up to 57.20) while maintaining a tight concentration in link-purpose errors (Criterion 2.4.4: Std = 1.88, IQR = 1.75). Finally, Community

C2 ($N = 55$) shows the highest overall error burden and the greatest internal heterogeneity. In this cluster, the pronounced divergence between parametric and non-parametric metrics indicates notable positive skewness and tail effects, particularly in Criterion 1.4.3, where the mean of 45.81 exceeds the median of 38.68 (Std = 29.86, Max = 78.60), and in Criterion 1.1.1, where the mean reaches 17.45 against a median of 11.45. This discrepancy suggests that although typical institutions within C2 already encounter substantial non-compliance levels, the group average is pulled upward by a subset of institutions exhibiting severe failure accumulation.

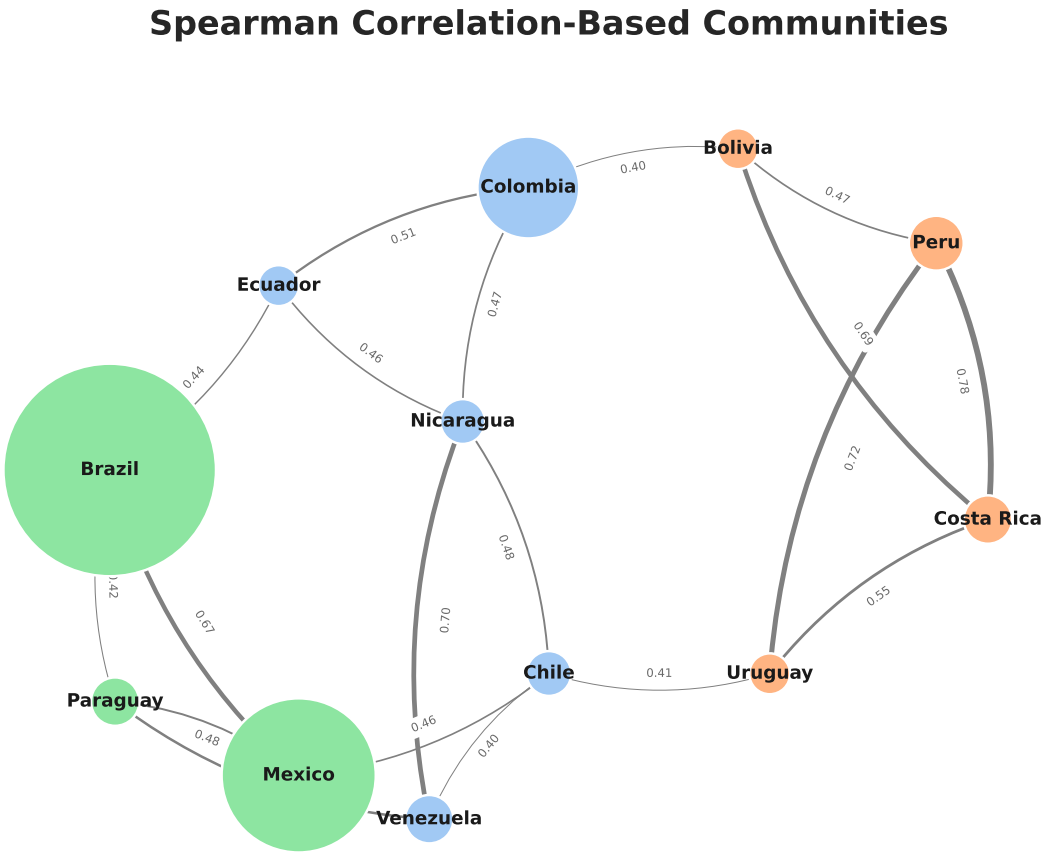


Figure 4. Country-level network of web accessibility error profiles in Latin America. Each node represents a country with at least five evaluated university websites (n = [number of countries] countries). An undirected edge connects two countries when the Spearman correlation between their mean frequencies across 13 WCAG error types is positive and $\rho \geq 0.40$; correlations below this threshold are not displayed. Node colors indicate communities detected by the Louvain algorithm, and edge widths represent the strength of the retained correlations.

Table 3. Descriptive statistics of the top WCAG error criteria across detected communities.

Community	Criterion	Mean	Std	Median	IQR	Min	Max
C0 ($N = 26$)	1.4.3	33.48	17.12	28.92	17.87	18.86	57.20
	2.4.4	10.67	1.88	10.98	1.75	8.14	12.60
	1.1.1	10.55	5.69	9.90	8.54	5.40	17.00
C1 ($N = 213$)	1.4.3	20.20	6.19	17.71	7.63	13.43	28.67
	2.4.4	14.16	3.95	12.92	1.91	10.57	20.83
	1.1.1	12.60	4.91	14.04	6.42	6.86	18.83
C2 ($N = 55$)	1.4.3	45.81	29.86	38.68	29.22	20.17	78.60
	2.4.4	20.75	5.14	18.20	4.64	17.39	26.67
	1.1.1	17.45	11.22	11.45	9.95	10.50	30.40

These findings support the hypothesis of a shared regional accessibility profile. The primary distinction among the analyzed countries lies not in the type of barriers present, but rather in their volume and dispersion. This consistency indicates that higher education institutions face a systemic technical baseline, where identical accessibility deficiencies recur across borders regardless of national or institutional context.

4.4. Descriptive Analysis of Accessibility Errors by WCAG Criterion

Table 4 summarizes the descriptive statistics for the 13 WCAG criteria that presented violations across the evaluated universities, showing that accessibility errors are concentrated in just three primary criteria that together account for 83.98% of all detected issues: Criterion 1.4.3 (Contrast Minimum) with 39.86% of the total (8,031 issues), followed by Criterion 2.4.4 (Link Purpose) with 22.40% (4,513 issues), and Criterion 1.1.1 (Non-text Content) with 21.72% (4,376 issues). In contrast, the remaining ten criteria collectively account for only 16.02% of the failures, with requirements such as 2.2.1 (Timing Adjustable) and 2.2.2 (Pause, Stop, Hide) registering relative frequencies near 0.00%.

Table 4. Descriptive statistics of detected accessibility errors by WCAG criterion.

WCAG Criterion	Total Freq.	Relative Freq. (%)	Universities with Errors	Mean	Std. Dev.	Median	IQR	Max
1.4.3 Contrast (Minimum) (Level AA)	8031	39.86	285	25.10	30.42	16.0	27.25	236.0
2.4.4 Link Purpose (In Context) (Level A)	4513	22.40	290	14.10	14.93	10.0	15.00	95.0
1.1.1 Non-text Content (Level A)	4376	21.72	288	13.68	20.01	8.0	15.25	180.0
1.3.1 Info and Relationships (Level A)	1099	5.45	200	3.43	8.58	1.0	4.00	129.0
2.4.6 Headings and Labels (Level AA)	812	4.03	186	2.54	4.21	1.0	3.00	30.0
3.3.2 Labels or Instructions (Level A)	656	3.26	164	2.05	3.74	1.0	3.00	30.0
4.1.2 Name, Role, Value (Level A)	376	1.87	39	1.18	7.81	0.0	0.00	129.0
2.4.1 Bypass Blocks (Level A)	157	0.78	54	0.49	2.03	0.0	0.00	26.0
2.1.1 Keyboard (Level A)	89	0.44	16	0.28	1.73	0.0	0.00	17.0
3.1.1 Language of Page (Level A)	30	0.15	29	0.09	0.30	0.0	0.00	2.0
2.4.2 Page Titled (Level A)	7	0.03	6	0.02	0.17	0.0	0.00	2.0
2.2.1 Timing Adjustable (Level A)	2	0.01	2	0.01	0.08	0.0	0.00	1.0
2.2.2 Pause, Stop, Hide (Level A)	1	<0.01	1	0.00	0.06	0.0	0.00	1.0

Regarding institutional reach, the data reveals that certain accessibility barriers are systemic across the evaluated sample ($N = 320$). In particular, three criteria affect almost the entire set of portals, each surpassing 280 institutions: Criterion 2.4.4 (Link Purpose) is present in 290 universities, followed by Criterion 1.1.1 (Non-text Content) with 288 universities, and Criterion 1.4.3 (Contrast Minimum) with 285 universities. In contrast, other criteria show very limited institutional spread, such as Criterion 2.2.1 (Timing Adjustable) and Criterion 2.2.2 (Pause, Stop, Hide), which were detected in only two universities and one university, respectively.

Finally, the central tendency and dispersion metrics reveal strong positive skewness across the evaluated criteria, as the mean consistently exceeds the median and the standard deviation often surpasses the mean itself. In addition, the interquartile range (IQR) shows that typical universities (the middle 50%) have a relatively narrow spread of errors, like 27.25 for Criterion 1.4.3 and 15.25 for Criterion 1.1.1. For the seven least frequent criteria, the IQR drops to 0.00, meaning that most portals had zero violations in those areas. In sharp contrast, the maximum values highlight extreme outlier behavior, reaching peaks of 236.0 in Criterion 1.4.3, 180.0 in Criterion 1.1.1, and 129.0 in both Criterion 1.3.1 and Criterion 4.1.2. This substantial distance between modest medians, bounded IQRs, and extreme maximums suggests that while the majority of universities maintain relatively low-to-moderate error

counts, a small group of poorly accessible portals accumulates massive error volumes, directly inflating the sample averages.

4.5. Typology of Errors

Figure 5 shows the distribution of errors by accessibility principle. The Perceivable principle is the one with the highest percentage (67%) followed by Operable (27.7%). The other two principles just represent 5.3% of the total. The results show that the majority of accessibility barriers affect users at the most fundamental interaction levels.

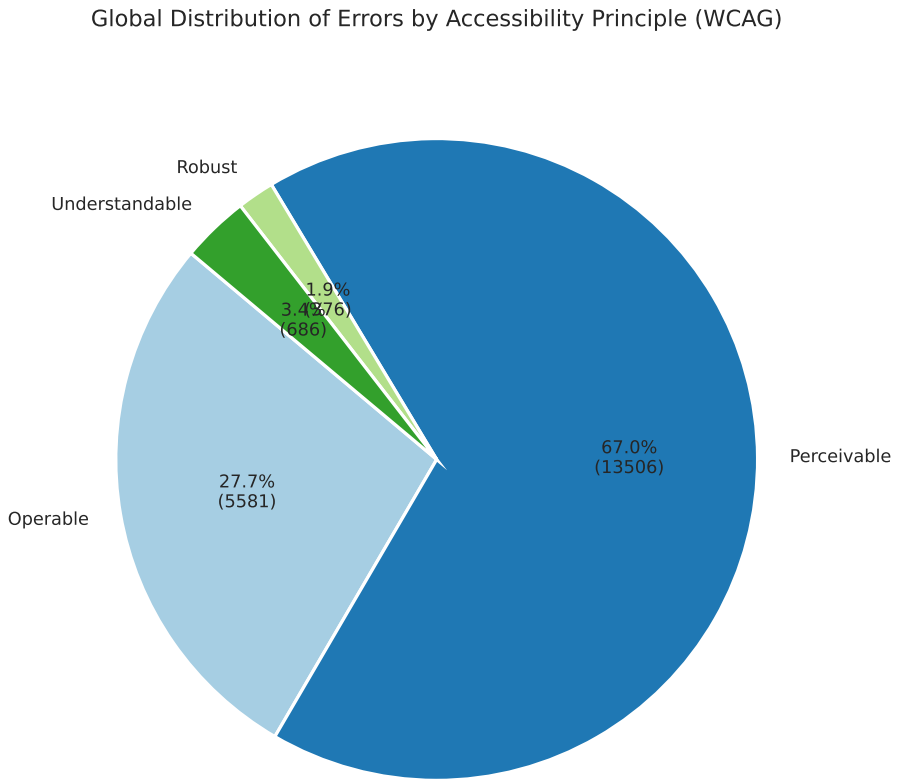


Figure 5. Percentage of errors by accessibility principle.

Figure 6 shows the distribution of errors considering the specific type of error and the accessibility principle to which it belongs. The most common error is 1.4.3 Contrast, which refers to failures in color contrast compliance. The second most frequent error is 2.4.4 Link Purpose about unclear clickable text. The next two errors: 1.1.1 Non-text Context and 1.3.1 Info and Relationship, are also part of the Perceivable principle, which indicates a major problem in ensuring that content can be properly interpreted by assistive technologies and perceived by users with visual impairments.

Figure 7 illustrates the percentage of universities failing to comply with each evaluated WCAG requirement, supporting the results observed previously. Nearly 90% of institutions present violations in criteria 2.4.4, 1.1.1, and 1.4.3, while three other criteria also affect more than half of the sample: 1.3.1 Info and Relationships (62.5%), 2.4.6 Headings and Labels (58.1%), and 3.3.2 Labels or Instructions (51.2%). These findings point to widespread difficulties with semantic content structuring, improper heading hierarchies, and inadequate instructions for interactive elements. Then, a substantial gap separates these six criteria from the rest of the catalog, as institutional prevalence plummets to just 16.9% for the next requirement. Given this highly skewed distribution, only the six criteria present in a majority (>50%) of the evaluated institutions were retained for the subsequent density-based clustering (DBSCAN) and similarity network analysis. Applying this prevalence threshold focuses the analysis on the most relevant accessibility barriers needed to accurately charac-

terize the Latin American ecosystem. Incorporating the remaining low-frequency errors is practically irrelevant for a macro-level diagnosis.

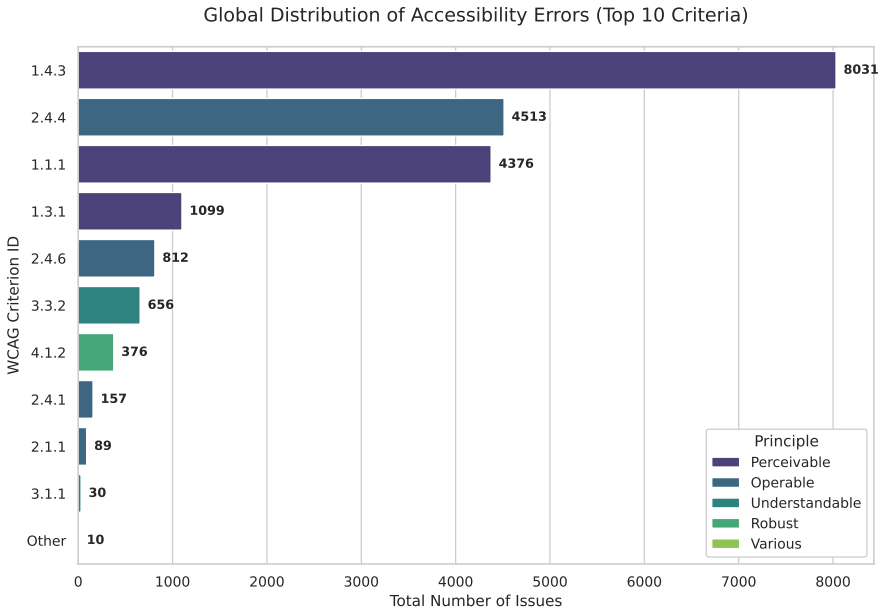


Figure 6. Most frequent type of errors.

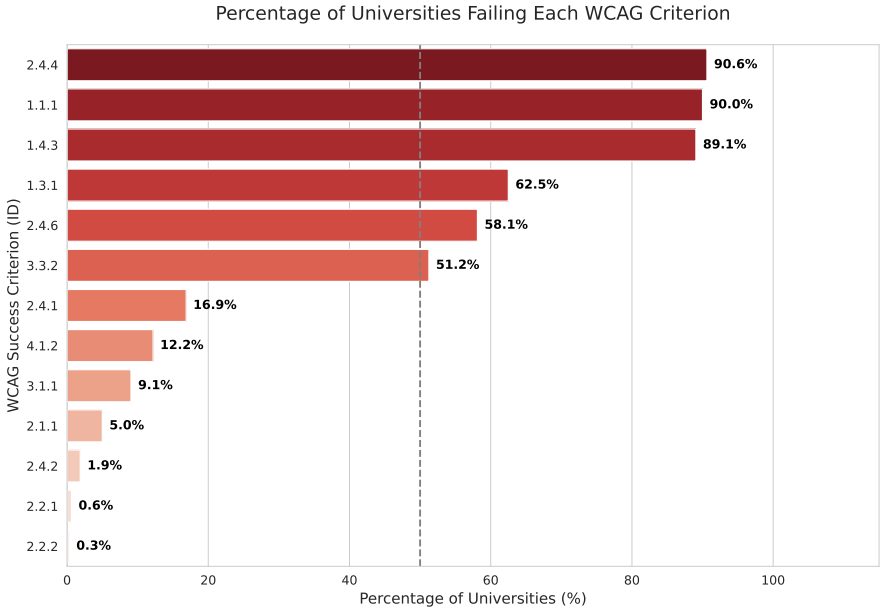


Figure 7. Percentage of universities that do not comply with the accessibility criteria. The different colors represent the various WCAG success criteria evaluated.

4.6. Detecting Accessibility Outliers

Prior to clustering, the six retained features were standardized using z-score normalization to ensure that differences in error scale did not bias the Euclidean distance calculations. The Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm was then applied to capture structural compliance patterns and detect institutional anomalies. It identifies clusters of arbitrary shapes without requiring an a priori specification of groups, while naturally isolating universities that deviate from regional behavior as noise points.

We configured the clustering modelh $\epsilon = 3.0$ and MinPts $= 12$. The MinPts parameter followed the standard heuristic of twice the dimensionality ($2 \times D$), corresponding to the six error criteria retained after applying the $> 50\%$ prevalence threshold. Meanwhile, ϵ was

selected using the elbow method on the *k*-distance graph, marking the point where density drops sharply. Under this setup, DBSCAN isolated 9 institutions as structural outliers, characterizing extreme systemic failures that deviate significantly from the main cluster.

Table 5 presents the comparative mean values between the mainstream regional group and the structural outliers across the primary criteria. While all evaluated dimensions exhibit a ratio greater than 2.0, the disparity in criterion 1.3.1 is particularly striking, with a 10.93-fold increase in error rates among outliers. This gap signals a profound lack of structural accessibility and suggests that these institutions fail to implement basic semantic labeling and information architecture.

Table 5. Comparison of mean accessibility errors between DBSCAN outliers and the mainstream cluster.

WCAG Success Criterion	Outliers (−1)	Mainstream (0)	Ratio (×)
1.3.1 Info and Relationships (Level A)	29.33	2.68	10.93
1.1.1 Non-text Content (Level A)	85.56	11.59	7.38
2.4.6 Headings and Labels (Level AA)	14.89	2.18	6.83
3.3.2 Labels or Instructions (Level A)	11.78	1.77	6.66
2.4.4 Link Purpose (Level A)	38.22	13.41	2.85
1.4.3 Contrast (Minimum) (Level AA)	52.11	24.32	2.14

Figure 8 presents a radar plot illustrating the disparities in mean error values between the two identified groups across all evaluated criteria. The visualization reveals that the area occupied by the mainstream group is merely a fraction of the vast area covered by the outliers. Notably, Criterion 1.1.1 emerges as the most critical failure point for the outlier group, reaching its highest absolute value. Furthermore, the distinct morphology (shape) of the two profiles indicates that the prevalence and prioritization of accessibility errors differ.

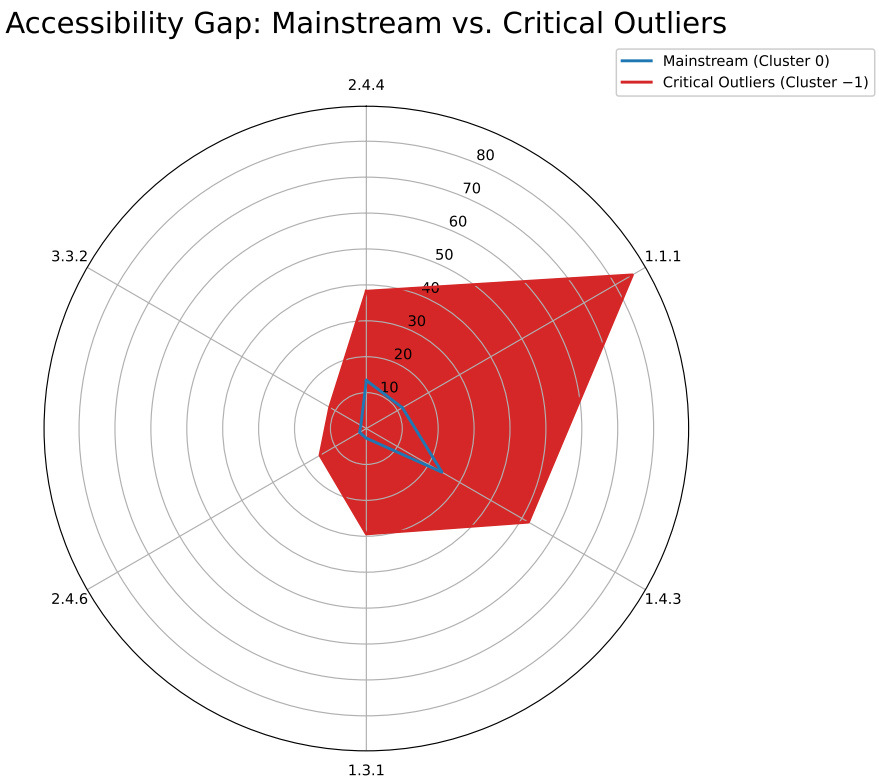


Figure 8. Percentage of universities that do not comply with the accessibility criteria.

4.7. Topological Analysis of Accessibility

We performed a Similarity Network Analysis (SNA) to examine the structural relationships between institutions. The network was constructed using the same six high-incidence dimensions from the previous clustering stage.

We calculated similarity between institutions using Cosine Similarity on these dimensions. A similarity threshold of $\tau = 0.85$ was selected to retain high-similarity relationships while filtering weak associations, thereby improving the interpretability of the resulting network. An undirected edge connected two university nodes when the cosine similarity between their vectors was ≥ 0.85, and the similarity value was used as the edge weight. Then, the Small-World coefficient (σ) was calculated to evaluate the integration of the similarity network. This analysis compares the observed network against a baseline Erdős-Rényi random graph with equivalent size and density.

The results show a Clustering Coefficient (C) of 0.6850, significantly higher than the random baseline ($C_{rand} = 0.1223$). Meanwhile, the Average Shortest Path Length (L) remains low at 4.01, comparable to the random expectation ($L_{rand} = 1.88$). These metrics yield a Small-World index (σ) of 2.63. According to the Watts-Strogatz model, a $\sigma > 1$ confirms the existence of a Small-World network.

This finding indicates high regional cohesion: Latin American universities are organized into dense, highly similar technical clusters, while simultaneously maintaining short 'distances' of similarity across the entire continent. This suggests that accessibility failures are not isolated incidents but part of a homogeneous technical ecosystem where specific patterns of digital exclusion are replicated and widespread throughout the region.

After that, community detection was conducted using the Louvain Method, a multilevel modularity optimization algorithm. This approach iteratively groups nodes to maximize Modularity (Q), a metric that compares the density of connections within communities against a null model.

The analysis yielded a modularity score of $Q = 0.3809$, indicating a non-trivial community structure within the accessibility similarity network.

Although 20 distinct communities were detected, a Pareto-based selection was used to focus on the most representative patterns. This approach identified five primary communities that constitute the core of the network and represent 81.56% of the total sample.

A heat map analysis (Figure 9) was performed using the mean error rates of the six high-prevalence criteria. Community 2 exhibits the most critical deficiencies in semantic architecture, leading the region in violations related to Information and Relationships (1.3.1), Headings and Labels (2.4.6), and Labels or Instructions (3.3.2). Community 18 shows a highly specialized pattern of failure, with a disproportionate average of 76.00 errors in Color Contrast (1.4.3), suggesting the use of shared, inaccessible design frameworks. Community 15 is defined by significant barriers in Link Purpose (2.4.4) and Non-text Content (1.1.1), indicating a lack of descriptive context for interactive elements and alternative text for media. Finally, communities 5 and 7 represent the most populous groups with comparatively lower error rates, although community 7 still maintains a notable peak in contrast issues.

These results show that a limited number of technical profiles account for most accessibility barriers across the region's educational websites. This concentration offers a strategic roadmap for intervention: by addressing the prominent errors identified in these core clusters, developers and policymakers can prioritize interventions for a potentially high systemic benefit addressing the most pervasive obstacles to digital inclusion across the continent.

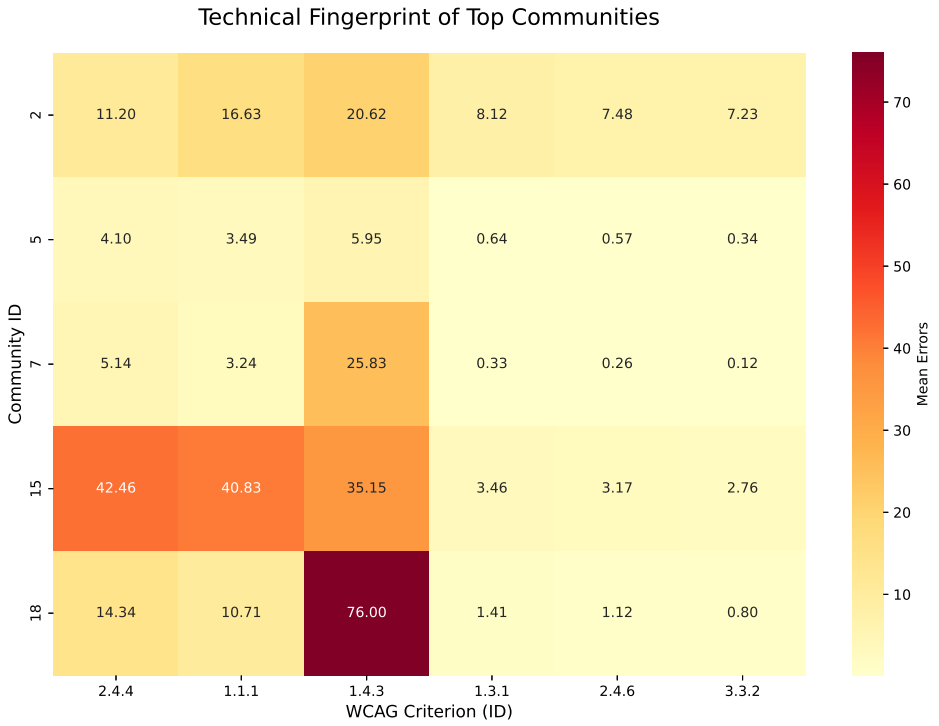


Figure 9. Heat map of the most representative profiles.

5. Proposed Data-Driven Web Accessibility Remediation Framework (DD-WARF)

The proposed DD-WARF transforms the empirical insights gathered from the preceding analyses into a formal, prescriptive architecture for accessibility diagnostics. An overview of the framework is shown in Figure 10.

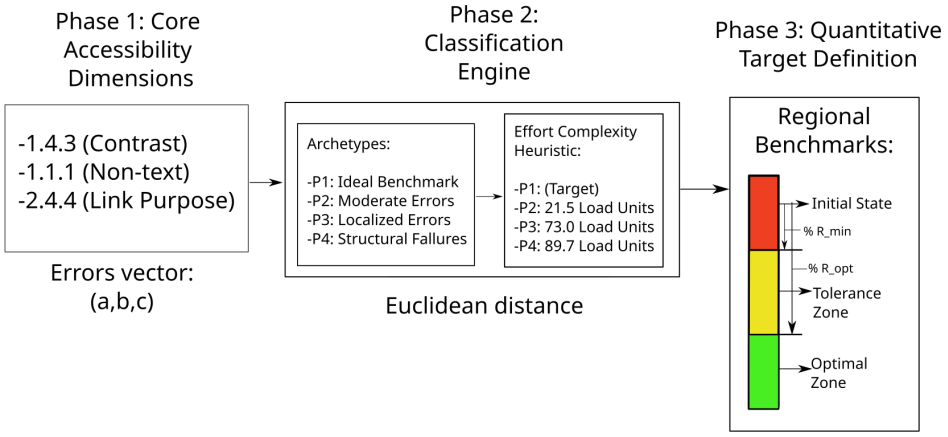


Figure 10. Framework overview. It illustrates the flow from the input (errors vector) to the final diagnosis that measures the distance between the current state and the recommended state.

DD-WARF operates in two stages. First, automated tools generate website-level observations of detectable accessibility errors. Second, the framework analyzes the distribution of those observations across institutions: DBSCAN identifies atypical profiles, similarity network analysis reveals recurring patterns, and the resulting profiles inform website classification. The Weighted Migration Effort Index (WMEI) then provides a relative estimate to support remediation prioritization. Thus, the framework's contribution lies in turning automated test outputs into comparative diagnosis and planning support. Its remediation-

impact estimates are simulated scenarios, not measurements of changes after repairs were implemented.

The framework reduces the high dimensionality of the WCAG guidelines to support actionable remediation planning. By introducing a taxonomic classification focused exclusively on 1.4.3 Contrast, 1.1.1 Non-text Content, and 2.4.4 Link Purpose, the model targets criteria that account for 83.98% of the regional error volume. Focusing on these common barriers allows decision-makers to prioritize technical interventions rather than trying to address all isolated violations at once based on the regional patterns identified during the community detection phase. Although the Louvain algorithm initially detected five primary groups representing roughly 80% of the regional ecosystem (Figure 9), communities 2 and 7 were merged into a single archetype because their mean error values were highly similar. These four resulting profiles serve as multi-dimensional centroids in an $\mathbb{R}^3$ space, allowing any given website to be classified and its remediation path to be calculated based on its proximity to the Ideal Reference (Profile 1).

5.1. Classification Strategy

To classify a specific website W, the framework represents its error density as a three-dimensional vector $\mathbf{v}_w = [e_{1.4.3}, e_{1.1.1}, e_{2.4.4}]$. The classification is then performed automatically by calculating the Euclidean distance between the website's vector and the centroids $\mathbf{C}_i$ of the four established archetypes (see Table 6).

Table 6. Consolidated remediation archetypes and their dimensional centroids.

Remediation Archetype	Contrast (1.4.3)	Non-Text Content (1.1.1)	Link Purpose (2.4.4)
Profile 1: Ideal	5.95	3.49	4.10
Profile 2: Medium Errors	22.76	11.12	8.71
Profile 3: Localized Error	76.00	10.71	14.34
Profile 4: Structural Errors	35.15	40.83	42.46

The website is assigned to the profile that minimizes the geometric distance, according to the following objective function:

$$\text{Profile}(W) = \arg\min_{i \in \{1,\ldots,4\}} \|\mathbf{v}_w - \mathbf{C}_i\|, \tag{1}$$

where the distance $\|\mathbf{v}_w - \mathbf{C}_i\|$ is defined as:

$$d(\mathbf{v}_w, \mathbf{C}_i) = \sqrt{\sum_{k \in \{1.4.3, 1.1.1, 2.4.4\}} (e_k - c_{i,k})^2}. \tag{2}$$

This automated approach allows developers and regional regulators to rapidly categorize institutional portals without requiring a full-scale manual audit, enabling real-time monitoring and large-scale benchmarking across the Latin American web ecosystem.

5.2. Weighted Migration Effort Index (WMEI)

To provide a pragmatic roadmap for remediation, the DD-WARF introduces the Weighted Migration Effort Index (WMEI). It functions as heuristic index of relative remediation complexity, not an estimate of engineering hours. Inspired by Agile estimation practices, the index assigns weights based on the Fibonacci sequence, where a higher numerical value reflects a greater technical effort and complexity required for remediation. This non-linear progression reflects the underlying technical domains of the violations:

- Contrast (1.4.3): $w = 1$. Typically involves low-effort, global CSS cosmetic adjustments.
- Non-text Content (1.1.1): $w = 2$. Requires localized content generation and HTML attribute modification.
- Link Purpose (2.4.4): $w = 3$. Often demands complex structural refactoring of the DOM and navigation logic.

The migration effort from a given profile centroid $\mathbf{C}_i$ to the Ideal Reference $\mathbf{C}_1$ is calculated using the Weighted Euclidean Distance:

$$WMEI(\mathbf{C}_i) = \sqrt{\sum_{k \in \{1,2,3\}} w_k (c_{i,k} - c_{1,k})^2} \quad (3)$$

Table 7 summarizes the calculated indices for the established archetypes. These values provide a relative benchmark for preliminary remediation planning.

Table 7. Calculated technical remediation effort by accessibility profile.

Remediation Archetype	Weighted Effort (Load Units)
Profile 4: Structural Errors	89.7
Profile 3: Localized Errors	73.0
Profile 2: Medium Errors	21.5

The WMEI facilitates preliminary technical planning at scale by providing a relative comparison of remediation complexity across accessibility profiles. By abstracting specific institutional constraints, such as developer headcount, work hours, or financial budgets, the index functions as a structural heuristic to prioritize interventions.

This metric allows decision-makers to quantify the "technical debt" of their institutional portals. While the WMEI provides a macro-level effort estimation, a more granular tool for calculating precise error-reduction targets will be detailed in the next section.

5.3. Actionable Remediation Roadmap by Architectural Profile

By interpreting the geometric distance as a proxy for systemic technical debt, we can formalize specific engineering roadmaps for each distinct architectural archetype.

5.3.1. Profile 2: Medium Errors

Representing the lowest technical friction across the sample, Profile 2 registers a minor effort footprint of WMEI = 21.5.

The diagnosis suggests that the structural components, global templates, and stylesheets appear fundamentally robust and compliant. The remaining defects tend to be behavioral, potentially resulting from human error during daily content operations, such as content editors omitting image alternative text or injecting non-compliant inline styles through rich-text interfaces.

Remediation should focus entirely on content governance and editorial training rather than code refactoring. Implementing structural constraints within the CMS WYSIWYG editors—such as making the `alt` attribute a mandatory fields prior to publishing—will quickly resolve these low-hanging fruits, driving the platform into full compliance.

5.3.2. Profile 3: Localized Error

Profile 3 presents an asymmetric statistical behavior, yielding a high cumulative index of WMEI = 73.0 that is driven almost exclusively by the 1.4.3 Contrast criterion (≈76.00).

This diagnoses a presentation-layer failure typically dictated by restrictive corporate brand identities, non-inclusive design languages, or misconfigured global CSS utility

frameworks. The remediation strategy for this archetype centers on a centralized refactoring of the presentation assets (CSS/SCSS).

Although the overall WMEI appears high, the remediation effort is highly leveraged; since the underlying issue is mapped to Contrast, which carries a lower agile weight ($w = 1$), altering global design tokens, updating theme variables, and enforcing a strict 4.5:1 luminance ratio will cause the geometric distance of this entire profile to collapse rapidly toward the reference profile with minimal code disruption.

5.3.3. Profile 4: Structural Errors

Characterized by a substantial technical burden of WMEI = 89.7, this profile shows severe, symmetrical degradation across all three vector dimensions.

The primary diagnosis for Profile 4 suggests broken or poorly engineered base templates within the Content Management System (CMS), a potential heavy reliance on non-compliant legacy dependencies, and a fundamental absence of native semantic HTML.

Development teams could prioritize refactoring the core backend output and global template wrappers. Furthermore, organizations should consider implementing governance gates, such as integrating automated accessibility compliance checkers directly into the continuous integration and continuous deployment (CI/CD) pipelines to help mitigate further regression.

5.4. Empirical Performance Envelopes and Remediation Equations

A common pitfall in web accessibility strategies is the pursuit of a "zero-error" state, which is often technically and financially unsustainable for large-scale institutional ecosystems. The DD-WARF shifts this paradigm by establishing Empirical Performance Envelopes based on the observed behavior of the high-performing cluster (Profile 1).

By analyzing the distribution of the "Ideal" community, we defined two reference benchmarks for each dimension:

- Recommended Target (T_{rec}): Defined by the median value, representing the performance of the top 50% of the reference group.
- Maximum Tolerance (T_{max}): Defined by the maximum value within the reference group, marking the boundary of baseline regional performance.

These thresholds, summarized in Table 8, allow the framework to calculate the recommended *Reduction Target* ($R\%$) for any given institution.

Table 8. Empirical remediation thresholds derived from the Profile 1 reference archetype.

WCAG Dimension	$T_{\mathbf{rec}}$ (Median)	$T_{\mathbf{max}}$ (Maximum)
WCAG 1.4.3 Contrast (Minimum)	5.0	20.0
WCAG 1.1.1 Non-text Content	2.0	15.0
WCAG 2.4.4 Link Purpose	3.0	13.0

where T_{rec} represents the median error incidence of the reference archetype and T_{max} represents the maximum observed error incidence within the same profile.

To automate the remediation roadmap, we propose the following equations to determine the minimum and recommended reduction percentages targeted for a website with an initial error count E_{actual}:

$$R_{min} = \max\left(0, \frac{E_{actual} - T_{max}}{E_{actual}}\right) \times 100 \tag{4}$$

$$R_{rec} = \max\left(0, \frac{E_{actual} - T_{rec}}{E_{actual}}\right) \times 100 \tag{5}$$

The inclusion of the $\max(0, \dots)$ function ensures that the framework recognizes sites already within the tolerance zone, yielding a 0% reduction target. This heuristic approach transforms a subjective audit into a set of engineering targets, allowing developers to prioritize interventions that move the system from a "Critical" state to a "Baseline" regional benchmark.

5.5. Ecosystem-Wide Classification Using Derived Archetypes

To describe how the derived archetypes categorize the institutional websites, the minimum Euclidean distance classification algorithm was applied to the same Latin American institutional dataset used to construct the archetypes.

The resulting assignments reflect the structural composition of this sample, serving as a descriptive baseline for the evaluated ecosystem.

Table 9 provides a descriptive overview of the structural maturity of the Latin American university web ecosystem. Profile 1 (Ideal) constitutes the largest individual share, encompassing 39.69% ($n = 127$) of the analyzed institutions.

Table 9. Distribution of Latin American universities across DD-WARF remediation archetypes

Remediation Archetype	Universities (*N*)	Percentage (%)
Profile 1: Ideal	127	39.69
Profile 2: Medium Errors	117	36.56
Profile 3: Localized Errors	37	11.56
Profile 4: Structural Errors	39	12.19
Total	320	100.00

This concentration indicates that achieving well-bounded accessibility thresholds may be a plausible goal given the technological and financial constraints of the region.

Furthermore, when combined with Profile 2 (Medium Errors), a significant 76.25% of the regional ecosystem falls within a low-to-moderate technical debt threshold, positioning the vast majority of institutions within a manageable mitigation distance.

Conversely, the remaining 23.75% of the dataset is split almost equally between severe architectural failure (Profile 4, 12.19%) and highly asymmetric non-compliance (Profile 3, 11.56%).

These two categories represent the high-priority intervention groups for regional policy. While institutions in Profile 3 require a highly localized, low-cost intervention (primarily style-sheet and contrast calibration), the portals categorized under Profile 4 demand systemic template refactoring.

To visually interpret the geometric boundaries, clustering density, and behavioral anomalies of these institutional profiles, a multi-panel scatter plot was constructed (Figure 11). The visualization maintains Criterion 1.1.1 as the fixed vertical baseline (Y-axis) across both panels, while alternating the horizontal reference (X-axis) to map Criterion 1.4.3 and Criterion 2.4.4, respectively.

In Panel A, a clear spatial separation is observed: Profile 1 (Ideal) tightly clusters near the origin, whereas Profile 3 (Localized) extends horizontally along the X-axis, indicating high contrast vulnerability but low non-text errors. Conversely, Profile 4 (Structural Errors) exhibits high density along the upper regions of the Y-axis.

In Panel B, Profiles 1, 2, and 3 show substantial geometric overlap near the origin. However, Profile 4 (Structural Errors) displays a distinct, positively sloped linear distribution, revealing a strong co-occurrence and apparent correlation between Link Purpose Errors and Non-text Content Errors within this specific cohort.

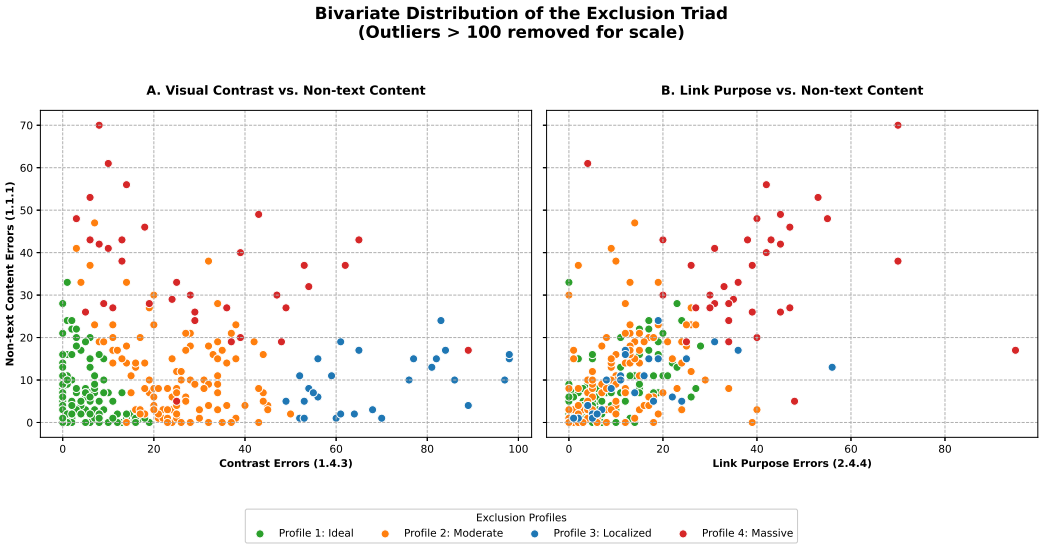


Figure 11. Dataset classification using the proposed framework.

5.6. Simulated Remediation Scenarios

A computational remediation scenario analysis was conducted to estimate the theoretical system-level outcomes of the proposed remediation equations.

The analysis simulated a scenario in which all institutions in the regional ecosystem applied the minimum and recommended remediation targets (R_{min} and R_{rec}) exclusively to the three dominant WCAG dimensions comprising the Accessibility Error Triad, while all other detected WCAG violations remained unchanged. This controlled scenario enabled the estimation of the potential reduction in both Triad-specific errors and the overall volume of automatically detected accessibility errors, thereby providing an analytical projection of the potential utility of prioritized remediation strategies.

As shown in Table 10, the baseline state of the regional ecosystem revealed a total volume of 20,149 errors, with the Triad accounting for 16,920 (83.9% of the total exclusions).

Table 10. Estimated systemic reduction of targeted remediation under minimum- and recommended-effort scenarios.

Metric	Minimum Effort (R_{min})	Recommended Effort (R_{rec})
Simulated Decrease in Triad Error Counts	7,538	14,228
Reduction in Triad (%)	44.6	84.1
Total Error Volume Reduction (%)	37.4	70.6

The simulation results illustrate a Pareto-like efficiency. By strategically addressing only three variables, representing low-to-medium technical complexity according to the WMEI, the framework projects a theoretical reduction in total detected errors between 37.4% and 70.6%.

By utilizing the DD-WARF to identify and mitigate dominant centroids, regional stakeholders can prioritize the mitigation of the vast majority of structural exclusion with a heuristically derived allocation of technical resources. This highlights the framework as a scalable and pragmatic decision-support tool for large-scale web accessibility governance.

5.7. AI-Assisted Accessibility Recommendation Architecture

Figure 12 show that DD-WARF enables the identification of accessibility archetypes and remediation priorities; its practical adoption can be further enhanced by integrating

Large Language Models (LLMs). Figure 10 illustrates the proposed AI-assisted recommendation architecture, in which DD-WARF-generated accessibility profiles are combined with WCAG knowledge to automatically produce contextualized remediation recommendations. Rather than replacing accessibility experts, the LLM acts as an intelligent decision-support component that transforms analytical results into actionable technical guidance for developers and institutional stakeholders.

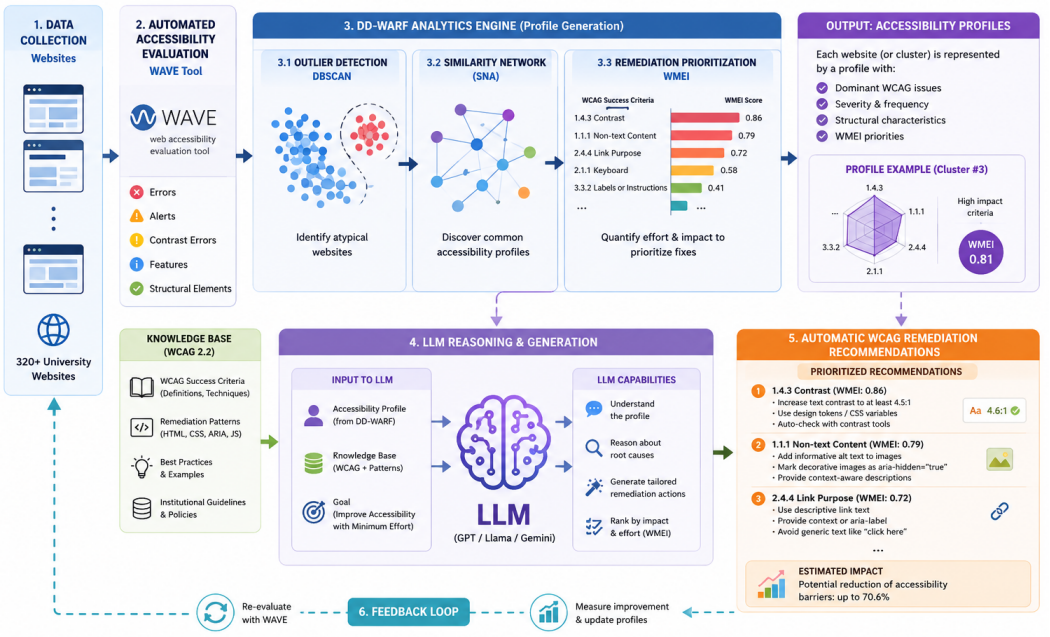


Figure 12. AI-Assisted Web Accessibility Remediation Architecture Based on DD-WARF and Large Language Models (LLMs).

The proposed architecture consists of six sequential stages. First, Lighthouse and WAVE evaluate institutional websites, and accessibility features are extracted. Second, DD-WARF performs statistical analysis, clustering, similarity network analysis, and WMEI estimation to generate accessibility profiles. Third, these profiles, together with WCAG 2.2 knowledge and remediation patterns, are provided as structured input to an LLM (e.g., GPT-4o, Llama 3.1, or Gemini 1.5 Flash). Finally, the LLM generates prioritized WCAG remediation recommendations, estimates their expected impact, and supports iterative accessibility improvement through continuous feedback.

6. Discussion

The regional consistency among Latin American universities suggests that generating a thousand different solutions is not necessary to guide initial remediation efforts. On the contrary, this uniformity allows for a prioritized roadmap where a few strategic interventions could target the majority of identified digital barriers. Specifically, the triad of exclusion, criteria 1.1.1, 1.4.3, and 2.4.4, represents the most pervasive obstacles. Addressing these three problems has the potential to facilitate the systemic improvement of the regional educational web ecosystem.

6.1. AI-Driven Remediation Strategies

For Criterion 1.1.1 (Non-text Content), the lack of alternative text remains a primary barrier for users of assistive technologies. Current State-of-the-Art (SOTA) multimodal models, such as GPT-4o, Gemini 1.5 Pro, or open-source alternatives like LLaVA (Large Language-and-Vision Assistant), can automatically analyze images and generate high-

quality, context-aware descriptive tags, ensuring that visual information is accessible to students with visual impairments [30–32].

Regarding Criterion 1.4.3 (Contrast Minimum), the failure often stems from aesthetic design choices that compromise legibility. This can be solved using algorithms that compute the relative luminance of the foreground and background colors. An automated script can also dynamically suggest or apply the nearest hexadecimal color value that meets the 4.5:1 threshold, preserving the institutional visual identity while ensuring readability [7,33]. Finally, for Criterion 2.4.4 (Link Purpose), the prevalence of generic labels such as "Read more" or "Download" creates navigational labyrinths for screen reader users. Open-source Large Language Models (LLMs) may be used to scan the DOM (Document Object Model) and analyze the surrounding text or parent elements of a link. By doing so, the AI can automatically rewrite the accessible name or the aria-label to provide a specific context, enhancing navigation.

6.2. Potential Implications for SDGs 4 and 10

These findings are consistent with recent studies reporting that most websites still present significant accessibility violations, particularly in contrast, alternative text, and navigation elements [11].

Addressing these accessibility barriers represents a decisive step toward achieving the United Nations Sustainable Development Goals (SDGs) [10] within the regional academic ecosystem. Addressing these accessibility barriers may contribute to SDG 4 (Quality Education) by supporting more inclusive digital learning environments. Moreover, the recurrence of similar accessibility patterns across institutions suggests that reusable remediation strategies could contribute to SDG 10 (Reduced Inequalities) by supporting more consistent accessibility practices across the regional higher-education ecosystem.

6.3. Responsible AI Considerations

The proposed AI-assisted accessibility remediation strategies are intended to support, rather than replace, expert human evaluation. Although Large Language Models (LLMs) and multimodal systems can automate the generation of alternative text, semantic labels, and accessibility recommendations, their outputs may contain inaccuracies, hallucinations, or contextually inappropriate descriptions [34].

To mitigate these risks, all AI-generated accessibility modifications should undergo human validation before deployment in production environments. In particular, accessibility corrections affecting users with disabilities must be carefully reviewed to ensure semantic accuracy, usability, and compliance with WCAG 2.2.

Furthermore, the framework promotes transparency and accountability by maintaining traceable remediation workflows and emphasizing human-in-the-loop validation processes. The proposed approach also seeks to minimize algorithmic bias by focusing on standardized accessibility criteria rather than user profiling or behavioral inference.

Finally, this study aligns with emerging Responsible AI principles that advocate for inclusive, transparent, and ethically governed AI systems in higher education digital ecosystems.

6.4. Limitations of the Study

The automated accessibility measures used in this study do not assess broader IS success dimensions, such as user satisfaction or social quality, nor do they measure stakeholder acceptance of AI-assisted recommendations. Evaluating these dimensions would require complementary studies with users, developers, and institutional decision-makers.

Despite the robustness of the analytical framework, several methodological limitations must be acknowledged. First, the evaluation relied exclusively on automated tools (WAVE [24] and Google Lighthouse [23]).

While effective for standardizing technical assessments across a large dataset, automated engines cannot detect issues requiring contextual, semantic, or cognitive interpretation. Because this study did not include manual audits by accessibility specialists or user-centered testing with individuals with disabilities, certain usability barriers inevitably remain unrecorded. The reported metrics reflect technical compliance rather than a guaranteed barrier-free user experience. The proposed 20-year update of the IS success model provides a broader lens for interpreting these findings. Automatically detected WCAG errors offer partial evidence about technical aspects of system quality and, for some criteria, information quality. They do not establish whether users can complete tasks successfully, understand the content, or access institutional services equitably, which relate to broader quality dimensions, including social quality. Accordingly, the error patterns and simulated reductions reported here should be interpreted as indicators for web accessibility diagnosis, not as observed improvements in user experience or overall IS success. Evaluating those outcomes requires expert inspection and direct studies with users with disabilities.

Second, the analysis was limited to institutional homepages, which were used as proxies for overall website accessibility. Although homepages serve as the primary gateway and generally reflect an institution's overarching design conventions, accessibility compliance levels may fluctuate significantly across internal pages, digital campus portals, or specific subdomains.

We derived the remediation archetypes from the analyzed dataset and used them to classify websites within that same dataset. Consequently, the distribution reported in Table 9 is descriptive and does not measure out-of-sample classification performance or the archetypes' transferability. External validation requires applying the predefined features, preprocessing steps, centroids, and assignment rule to an independent set of university websites without re-estimating the archetypes from those websites.

Although Sections 4.2 and 4.3 compare accessibility errors across countries, the study did not classify websites by page language or systematically code their design conventions, such as navigation patterns, templates, content-management systems, or component libraries. Country-level similarities and differences therefore cannot be attributed to language or design practices. Future studies should record these characteristics at the website level and examine their association with specific WCAG error types, while accounting for differences in country sample sizes.

Third, the simulated decrease in automatically detected error counts (37.4% to 70.6% reduction) represents a theoretical projection derived from a computational simulation. Future empirical studies are required to assess the practical challenges of implementing these targeted fixes in live university environments.

Finally, web accessibility is inherently dynamic. Due to continuous content updates, template migrations, and redesigns, the reported metrics represent specific snapshots in time of the higher education ecosystem rather than permanent conditions.

7. Conclusions

This study provides a large-scale, data-driven characterization of web accessibility barriers across 320 Latin American university websites and introduces DD-WARF, a computational framework for transforming automated accessibility findings into actionable remediation priorities. Rather than considering accessibility violations as isolated technical issues, the proposed approach integrates statistical analysis, anomaly detection, similarity network analysis, remediation archetypes, and effort estimation to characterize recurring accessibility patterns at both institutional and regional levels.

The statistical analysis indicated no significant association between institutional ranking and the observed accessibility error burden. Likewise, despite variations in absolute

error counts, the country-level analysis did not yield statistically significant differences in the overall distribution of accessibility errors. These findings suggest that the accessibility barriers identified in the dataset are not restricted to particular institutional rankings or national contexts.

A recurrent Accessibility Error Triad emerged from the analysis, comprising WCAG criteria 1.4.3 (Contrast (Minimum)), 1.1.1 (Non-text Content), and 2.4.4 (Link Purpose (In Context)). These criteria affected nearly 90% of the evaluated institutions and accounted for a substantial proportion of the automatically detected accessibility errors. A second group of recurrent barriers, including criteria 1.3.1 (Info and Relationships), 2.4.6 (Headings and Labels), and 3.3.2 (Labels or Instructions), affected approximately 50–60% of the analyzed websites, further illustrating the persistence of structural accessibility problems across the dataset.

Using these six WCAG dimensions, DBSCAN identified nine institutions with atypical accessibility profiles. For the analyzed criteria, these outliers exhibited mean error incidences between 2.14 and 10.93 times those observed in the mainstream cluster. Complementarily, cosine-similarity network analysis identified communities composed of institutions from different countries that shared similar accessibility profiles. Five communities concentrated 81.56% of the institutions, suggesting that a relatively small number of recurring technical patterns characterize a large proportion of the analyzed ecosystem. Among these patterns, WCAG 1.4.3 emerged as a particularly prominent barrier.

The network analysis also highlighted high connectivity and clustering among institutional accessibility profiles. These results suggest that many accessibility problems are shared across institutions rather than being entirely institution-specific. From a practical perspective, this commonality creates an opportunity to develop reusable remediation strategies that can be adapted across institutions exhibiting similar accessibility profiles, potentially reducing the cost and complexity of large-scale accessibility interventions.

Based on these empirical patterns, DD-WARF classifies institutional websites into four remediation archetypes: Ideal, Medium Errors, Localized Errors, and Structural Errors. The framework further maps the three dominant WCAG dimensions into a three-dimensional remediation space and introduces the Weighted Migration Effort Index (WMEI) to provide a relative estimate of the technical effort associated with moving each archetype toward the Ideal reference profile. WMEI is intended as a decision-support mechanism for prioritizing remediation and resource allocation rather than as an absolute measure of development time or cost.

Because achieving an absolute zero-error state may be operationally unrealistic for large and continuously evolving web environments, DD-WARF additionally incorporates empirically derived remediation thresholds. The minimum and recommended remediation targets (R_{min} and R_{rec}) provide alternative reference scenarios for reducing the dominant accessibility barriers. Under the simulated remediation scenarios, applying these targets to the Accessibility Error Triad resulted in an estimated reduction of 37.4–70.6% in the total volume of automatically detected accessibility errors. These estimates should be interpreted as potential remediation outcomes under the assumptions of the simulation rather than as observed improvements resulting from real-world interventions.

The mathematical formulation of the WMEI also relies on a heuristic weighting system. The summation and the assigned weights serve strictly as ordinal proxies for technical complexity and do not predict actual engineering hours, financial costs, or precise developer effort. Because real-world remediation depends heavily on institutional variables (such as legacy infrastructure, CMS constraints, and developer expertise), the current WMEI provides a theoretical baseline for prioritization. Future research could refine this model by calibrating the weights against empirical data from actual accessibility remediation projects.

The processed dataset is publicly available through Mendeley Data [11]. The manuscript also reports key analytical settings, including the DBSCAN parameters and network similarity threshold. Nevertheless, public access to the dataset alone does not ensure independent reproduction of the complete DD-WARF workflow. Reproducing every result would additionally require the analysis code, software versions, preprocessing steps, and a complete record of parameter settings. Until these materials are made available together, the framework should be regarded as documented and open to further verification, rather than as fully reproducible. Future research should extend the dataset to improve the representation of countries with smaller samples and evaluate the temporal stability of the identified accessibility archetypes. Longitudinal studies could determine whether the remediation profiles and network structures remain stable as institutional websites evolve and could measure the actual impact of accessibility policies and technical interventions. Further research should also empirically evaluate the proposed integration of large language models and multimodal models for generating context-aware remediation recommendations, with particular attention to accuracy, explainability, hallucination risks, and human oversight.

Finally, automated evaluation cannot capture all semantic, contextual, cognitive, and experiential accessibility barriers. Future validation should therefore combine automated assessment with expert inspection and direct evaluation involving users with disabilities. Such human-centered validation is essential for determining whether reductions in automatically detected WCAG violations translate into meaningful improvements in real-world accessibility. The public availability of the DD-WARF Accessibility Dataset supports reproducibility, benchmarking, and further research on scalable, data-driven approaches to web accessibility assessment and remediation.

Author Contributions: Conceptualization, P.A.-V. and M.G.-R.; methodology, P.A.-V.; software, M.G.-R., P.A.-V. and D.D.-B.; validation, D.D.-B., S.N.-V., M.G.-R., L.S.-U. and B.S.-A.; formal analysis, D.D.-B., S.N.-V., M.G.-R., L.S.-U. and B.S.-A.; investigation, D.D.-B., S.N.-V., M.G.-R., L.S.-U. and B.S.-A.; resources, P.A.-V.; data curation, P.A.-V., D.D.-B. and S.N.-V.; writing—original draft preparation, D.D.-B., S.N.-V., M.G.-R., L.S.-U. and B.S.-A.; writing—review and editing, D.D.-B., S.N.-V., M.G.-R., L.S.-U. and B.S.-A.; visualization, M.G.-R. and L.S.-U.; supervision, P.A.-V.; project administration, P.A.-V.; funding acquisition, P.A.-V. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Universidad de Las Américas-Ecuador as part of the internal research project 518.A.XV.24.

Data Availability Statement: The dataset analyzed during this study is publicly available at [11].

Acknowledgments: The authors used ChatGPT (OpenAI, version 5.5) and Grammarly version 6.8.263 exclusively to support grammar correction, language editing, readability enhancement, and proofreading of the manuscript. These tools were not used to generate scientific content, conduct analyses, interpret results, formulate conclusions, or make research decisions. The authors performed and validated all aspects of the research, including conceptualization, methodology, data collection, data analysis, interpretation of findings, and preparation of the final manuscript. The authors carefully reviewed and edited all AI-assisted suggestions and take full responsibility for the accuracy, originality, and integrity of the content presented in this work.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. W3C. *Accessibility Guidelines (WCAG) 3.0*; W3C Working Draft, 10 September 2026. Available online: https://www.w3.org/TR/2026/WD-wcag-3.0-20260910/ (accessed on 4 October 2026).
2. Naser, M.Z. A review of machine learning with small and limited data. *J. Big Data* **2026**, *13*, 18. https://doi.org/10.1186/s40537-025-01346-9.

3. W3C. *Web Content Accessibility Guidelines (WCAG) 2.0*; W3C Recommendation, 11 December 2008. Available online: https://www.w3.org/TR/2008/REC-WCAG20-20081211/ (accessed on 4 October 2026).
4. Zowghi, D.; Bano, M. AI for all: Diversity and Inclusion in AI. *AI Ethics* **2024**, *4*, 873–876. https://doi.org/10.1007/s43681-024-00485-8.
5. Acosta-Vargas, P.; Novillo-Villegas, S.; Salvador-Acosta, B.; Calvopina, M.; Kyriakidis, N.; Ortiz-Prado, E.; Salvador-Ullauri, L. Accessibility Analysis of Worldwide COVID-19-Related Information Portals. *Int. J. Environ. Res. Public Health* **2022**, *19*, 12102. https://doi.org/10.3390/ijerph191912102.
6. Drljić, K.; Konrad, S.Č.; Rutar, S.; Štemberger, T. Digital Equity and Sustainability in Higher Education. *Sustainability* **2025**, *17*, 2011. https://doi.org/10.3390/su17052011.
7. W3C. *Web Content Accessibility Guidelines (WCAG) 2.2*; World Wide Web Consortium Inc.: Wakefield, MA, USA, 2023.
8. Preiser, W.; Smith, K. *Universal Design Handbook*; McGraw-Hill Construction Media: New York, NY, USA, 2011; p. 492.
9. Lauer-Schmaltz, M.W.; Cash, P.; Rivera, D.G.T. ETHICA: Designing Human Digital Twins—A Systematic Review and Proposed Methodology. *IEEE Access* **2024**, *12*, 86947–86973. https://doi.org/10.1109/ACCESS.2024.3416517.
10. United Nations. The 17 Goals. 2026. Available online: https://sdgs.un.org/goals (accessed on: 4 October 2026).
11. Acosta-Vargas, P. *DD-WARF Accessibility Dataset*; Mendeley Data: London, UK, 2026; Version 1. https://doi.org/10.17632/8p7c7wjb5v.1.
12. Johnson, P.; Lilley, M. Software Prototype for the Ensemble of Automated Accessibility Evaluation Tools. *Commun. Comput. Inf. Sci.* **2022**, *1580 CCIS*, 532–539. https://doi.org/10.1007/978-3-031-06417-3_71.
13. Ara, J.; Sik-Lanyi, C.; Kelemen, A.; Guzsvinecz, T. An inclusive framework for automated web content accessibility evaluation. *Univers. Access Inf. Soc.* **2025**, *24*, 1581–1607. https://doi.org/10.1007/s10209-024-01164-5.
14. Abu Doush, I.; Sultan, K.; Al-Betar, M.A.; Almeraj, Z.; Alyasseri, Z.A.A.; Awadallah, M.A. Web Accessibility Automatic Evaluation Tools: To What Extent Can They Be Automated? *CCF Trans. Pervasive Comput. Interact.* **2023**, *5*, 288–320. https://doi.org/10.1007/s42486-023-00127-8.
15. Martins, B.; Duarte, C. A large-scale web accessibility analysis considering technology adoption. *Univers. Access Inf. Soc.* **2024**, *23*, 1857–1872. https://doi.org/10.1007/s10209-023-01010-0.
16. MacAkoglu, S.S.; Peker, S. Web accessibility performance analysis using web content accessibility guidelines and automated tools: A systematic literature review. In *Proceedings of the 2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA 2022), Ankara, Turkey, 9–11 June 2022*; IEEE: New York, NY, USA, 2022. https://doi.org/10.1109/HORA55278.2022.9799981.
17. Duarte, C.; Matos, I.; Carriço, L. Semantic content analysis supporting web accessibility evaluation. In Proceedings of the 15th International Web for All Conference, Lyon, France, 23–25 April 2018. https://doi.org/10.1145/3192714.3196828.
18. Gorrini, A.; Choubassi, R.; Rezaallah, A.; Presicce, D.; Boratto, L.; Laniado, D.; Aragón, P. Assessing the Level of Accessibility of Railway Public Transport for Women Passengers Using Location-Based Data: The Case of H2020 DIAMOND Project. In *Green Energy and Technology*; Springer: Cham, Switzerland, 2021; pp. 199–212. https://doi.org/10.1007/978-3-030-57332-4_14.
19. López-Gil, J.M.; Pereira, J. Turning manual web accessibility success criteria into automatic: An LLM-based approach. *Univers. Access Inf. Soc.* **2025**, *24*, 837–852. https://doi.org/10.1007/s10209-024-01108-z.
20. Palarimath, S.; Radhakrishnan, V.; Veerasamy, D.; Valsalan, P.; Viswan, V.; Thumu, M.B. Artificial Intelligence and Machine Learning for Accessibility: Smart Reader as an Assistive Tool for the Visually Impaired. In *Proceedings of the 2025 International Conference on Machine Learning and Autonomous Systems (ICMLAS), Prawet, Thailand, 10–12 March 2025*; IEEE: New York, NY, USA, 2025. pp. 424–430. https://doi.org/10.1109/ICMLAS64557.2025.10968797.
21. Kumar, N.; Wei, X.; Zhang, H. Agentic artificial intelligence as a new frontier in information systems: Promise, peril, and research opportunities. *Inf. Manag.* **2026**, *63*, 104317.
22. Tasleem, N.; Ansari, M.N.; Afzal, M.N.I. AI-Driven decision intelligence: A Bayesian framework for valuing HR technology readiness and international investment under uncertainty. *J. Model. Manag.* **2026**, 1–20. https://doi.org/10.1108/JM2-11-2025-0604.
23. Google. Lighthouse: Overview. 2026. Available online: https://developer.chrome.com/docs/lighthouse/overview/ (accessed on 23 August 2026).
24. WebAIM. WAVE Web Accessibility Evaluation Tool. 2026. Available online: https://wave.webaim.org/ (accessed on 4 October 2026).
25. Spearman, C. The proof and measurement of association between two things. *Am. J. Psychol.* **1904**, *15*, 72–101.
26. Ester, M.; Kriegel, H.P.; Sander, J.; Xu, X. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining (KDD'96), Portland, OR, USA, 2–4 August 1996*; AAAI Press: Washington, DC, USA, 1996; pp. 226–231.
27. Schubert, E.; Sander, J.; Ester, M.; Kriegel, H.P.; Xu, X. DBSCAN Revisited, Revisited: Why and How You Should (Still) Use DBSCAN. *ACM Trans. Database Syst.* **2017**, *42*, 1–21. https://doi.org/10.1145/3068335.

28. Cha, S.H. Comprehensive survey on distance/similarity measures between probability density functions. *Int. J. Math. Model. Methods Appl. Sci.* **2007**, *1*, 300–307.
29. Barabási, A.L. *Network Science*; Cambridge University Press: Cambridge, UK, 2016.
30. OpenAI. GPT-4 Technical Report. *arXiv* **2024**, arXiv:2303.08774. https://doi.org/10.48550/arXiv.2303.08774.
31. Google DeepMind. Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context. *arXiv* **2024**, arXiv:2403.05530.
32. Liu, H.; Li, C.; Wu, Q.; Lee, Y.J. Visual Instruction Tuning. *arXiv* **2023**, arXiv:2304.08485. https://doi.org/10.48550/arXiv.2304.08485.
33. Lalitha, A.R. Perceptually-Minimal Color Optimization for Web Accessibility: A Multi-Phase Constrained Approach. *arXiv* **2025**, arXiv:2512.05067. https://doi.org/10.48550/arXiv.2512.05067.
34. Baeza-Yates, R.; Fayyad, U.M. Responsible AI: An Urgent Mandate. *IEEE Intell. Syst.* **2024**, *39*, 12–17. https://doi.org/10.1109/MIS.2023.3343488.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.