informatics

MDPI

Article

Machine Learning and Generative AI in Administrative Processes in Peru: Administrative Efficiency in the National Public Sector

Miluska Odely Rodriguez Saavedra [1,*], Juliana Mery Bautista Lopez [1], Wilian Quispe Nina [1], Antonio Víctor Morales Gonzales [2], Iván Cuentas Galindo [2], Luis Miguel Campos Ascuña [2], Anthony Stefano Saenz Colana [2], Robinson Bernardino Almanza Cabe [3], Paola Gabriela Lujan Tito [4] and Sharon Veronika Liendo Teran [2]

[1] Faculty of Accounting and Financial Sciences, Graduate School, Universidad Nacional de San Agustín de Arequipa, Arequipa 04000, Peru; jbautistal@unsa.edu.pe (J.M.B.L.); wquispeni@unsa.edu.pe (W.Q.N.)
[2] Faculty of Economics, Universidad Nacional de San Agustín de Arequipa, Arequipa 04000, Peru; amoralesg@unsa.edu.pe (A.V.M.G.); icuentas@unsa.edu.pe (I.C.G.); lcamposas@unsa.edu.pe (L.M.C.A.); asaenzc@unsa.edu.pe (A.S.S.C.); sliendo@unsa.edu.pe (S.V.L.T.)
[3] Faculty of Social Sciences, Universidad Nacional de Moquegua, Moquegua 18001, Peru; ralmanzac@unam.edu.pe
[4] Faculty of Administration and Business, Universidad Tecnológica del Perú, Lima 15073, Peru; C28388@utp.edu.pe
* Correspondence: mrodriguezsaa@unsa.edu.pe

Abstract

Public organizations in Peru have committed substantial resources to artificial intelligence over recent years, yet evidence on whether these investments produce measurable returns has remained scarce. This study evaluated the causal impact of AI adoption on administrative efficiency across 20 Peruvian national public organizations, using a quasi-experimental design combining Difference-in-Differences with Propensity Score Matching, complemented by XGBoost version 1.7.6, Random Forest, GPT-4, and SHAP explainability analysis. The sample comprised 428 civil servants across treatment and control organizations. Results showed significant efficiency gains as perceived by civil servants through validated Likert instruments: work absenteeism decreased by 9.4%, processing times by 8.7%, and administrative costs by 18.2%, all at $p < 0.001$ with Cohen's d ranging from 0.55 to 0.90. The convergence between DiD and PSM estimates supports a causal reading of these effects. Four of five hypotheses were supported. AI delivered comparable efficiency gains regardless of institutional complexity, so H2 was not confirmed. Digital infrastructure significantly moderated AI effectiveness (H3: r = 0.198, $p = 0.004$). Higher resistance to change was significantly associated with lower efficiency outcomes (H5: r = −0.256, $p < 0.001$), reinforcing the role of proactive change management as a positive moderator of AI effectiveness. SHAP analysis revealed that training investment, specialized IT personnel, and resistance management together explained 51% of predictive importance, outweighing structural variables such as budget size or geographic location. These findings provide the first systematic causal evidence on AI efficiency in Peruvian public administration and offer actionable benchmarks for comparable middle-income public sectors.

Keywords: artificial intelligence; public sector; administrative efficiency; machine learning; model explainability; causal inference

Academic Editor: Manuel Pedro Rodríguez Bolívar

Received: 9 January 2026
Revised: 15 March 2026
Accepted: 16 March 2026
Published: 19 March 2026

1. Introduction

The relationship between technology and public administration does not follow a straight line. Over the past two decades, governments around the world have poured resources into digital systems, pushed by growing citizen demand for faster and more transparent services [1]. Artificial intelligence and machine learning moved to the center of this agenda. Advocates pointed to their potential to reshape how administrative data is handled, how decisions get made, and how institutions engage with the public [2]. That optimism, however, sits alongside well-documented concerns about algorithmic opacity, uneven access to infrastructure, and the real possibility that AI deployments deepen rather than narrow existing inequalities.

Comparative work from Europe and Asia offers useful reference points, even though Latin America is the primary setting of this analysis. Those studies show that AI adoption hinges not only on technological readiness but also on bureaucratic flexibility, digital literacy, and the capacity of institutions to adapt [3]. They also show that institutional and cultural barriers tend to matter more than the technology itself [4].

Across Latin America, local governments have moved from planning to implementation at varying speeds. Brazilian municipalities are deploying chatbots for citizen services and automated classification of tax inquiries [5]. Chilean regional governments use predictive analytics to manage public transportation. Colombian programs pilot machine learning for early identification of school dropout risk, while Mexican agencies have introduced generative AI to handle inquiries about administrative procedures [6]. The diversity of these cases points to how much institutional and territorial setting shape adoption trajectories.

Peru constitutes a well-defined stress-test environment for evaluating algorithmic performance under the infrastructure and data quality conditions that characterize a large portion of the world's public administrations. Broadband penetration in Metropolitan Lima reaches levels comparable to Southern European cities, while Amazonian regions record connectivity rates closer to sub-Saharan African baselines, producing a broadband gap of approximately 4.3 index points between the two extremes within a single national legal framework. That internal contrast allows the study to measure how the same algorithms perform across connectivity gradients without the confounding introduced by comparing across countries with different regulatory systems, budget structures, or data collection standards. The 20 national public organizations in the sample, spanning judicial bodies, electoral institutions, and central ministries, operate under a unified national regulatory framework for public administration [7]. (DS No. 157-2021-PCM) [8], which standardizes administrative procedures and digitization requirements across all national public entities. No prior study has evaluated ML performance across this set of national entities within a single regulatory environment. These characteristics make the findings directly transferable to public administrations in Indonesia, the Philippines, Colombia, South Africa, and other middle-income countries whose public sectors operate under analogous conditions of fragmented digital infrastructure, heterogeneous institutional capacity, and incomplete administrative digitization. The study covers the period January 2022 to December 2024. AI implementation in treatment organizations took place between 2022 and 2024. The survey instrument was administered to 428 civil servants across 20 national public organizations between January and April 2025, capturing an early AI adoption phase in which organizational and computational constraints are most consequential for deployment decisions.

From a methodological standpoint, this work contributes to informatics by providing the first systematic evaluation of ML performance under the kind of structural data heterogeneity that characterizes administrative systems in emerging economies. It also documents how LLM deployment behaves under authentic institutional pressures and assesses the

stability of explainability methods across organizationally diverse environments [9]. Those contributions carry value for countries in comparable situations globally.

The literature confirms AI's potential in the public sector, but it also documents the risks, among them opacity, access inequality, oversimplified decision-making, and the sidelining of vulnerable populations. Maragno et al. [10] draw attention to the conflicting ways algorithmic technologies play out in governance. Bian and Wang [11] finds that internal organizational factors are the primary bottleneck in successful implementations.

Despite its recent growth, the literature on AI in the public sector leaves three documented gaps [12]. The first is an overconcentration on national government strategies, leaving organizational-level dynamics underexplored, even though this is where administrative services and citizens directly interact. The second is a shortage of measurement frameworks capable of capturing AI integration across multiple administrative dimensions at once [13]. The third is a tendency toward one-dimensional analytical designs that treat local organizational factors and regional structural conditions as independent rather than interacting [14]. The combined effect is a study that captures neither micro-level institutional dynamics nor macro-level contextual effects with adequate precision.

The immediate motivation for this study was a practical problem in Peru. Public organizations committed substantial budgetary resources to digitization between 2022 and 2024, with no systematic evidence on whether that spending produced measurable returns in absenteeism, processing times, or operating costs. Without reliable estimates, policymakers face a choice between scaling programs of unverified effectiveness and abandoning tools that may genuinely improve performance. Neither position is defensible. A parallel gap concerns causality. Studies on AI in government have grown in number, but most rely on cross-sectional surveys or before–after comparisons that cannot establish whether AI caused the observed changes or whether other concurrent factors did [15]. Which specific organizational factors determine implementation success, and whether those factors differ by institution type or geographic region, remains unanswered.

From a computational standpoint, this study examines three problems that standard benchmark evaluations cannot replicate [16]. First, XGBoost, Random Forest, and GPT-4 are tested under administrative data conditions that violate the assumptions of controlled settings, specifically missing values, clustering across 20 national public organizations, and temporal autocorrelation. GPT-4 is evaluated as a deployed system, measuring prompt sensitivity and reproducibility across independent sessions. SHAP-based explainability is subjected to ordinal stability testing and cross-validated against LIME and permutation importance in an institutionally diverse setting where such validation has not previously been reported. These three evaluations, taken together, provide the informatics community with documented evidence of how standard algorithms perform when structural assumptions break down in practice [17].

Deploying machine learning in public sector environments turns out to be a different undertaking than deploying it in a research lab. Data that governments collect was never designed with algorithmic analysis in mind, and records accumulated over years of inconsistent administrative practice carry noise, gaps, and structural irregularities that curated benchmark datasets simply do not have [18]. Administrative processes also unfold over time in ways that create dependencies between observations, which quietly undermine the independence assumptions that standard models take for granted. Public accountability adds yet another layer that private sector deployments rarely face, since institutions using algorithms to influence decisions about citizen services carry a genuine obligation to explain what the model is doing in terms that non-technical stakeholders can evaluate. Generative AI operating in official capacities requires governance protocols that go well beyond accuracy measurement. This study contributes to the informatics community a

documented account of how these challenges play out in practice, bringing together ensemble methods for performance stability under uneven data quality, transformer-based models for unstructured administrative text, and additive feature attribution methods for the kind of transparent explanation that regulatory requirements demand [19]. The transferability of these findings extends to any national public administration where data quality constraints, multi-institutional clustering, and limited computational infrastructure define the deployment baseline, a profile that describes the majority of middle-income countries in Asia, Africa, and Latin America.

The four research questions are:

RQ1: To what extent does the implementation of artificial intelligence reduce work absenteeism, processing times, and administrative costs in Peruvian public organizations?

RQ2: How does this impact differ depending on the type of public organization and the geographical region of operation?

RQ3: Which organizational factors, including IT staff capacity, resistance to change, and training investment, moderate the effectiveness of AI implementation in public organizations?

RQ4: What patterns of heterogeneity of effects can be identified to inform differentiated digital transformation strategies in the Peruvian public sector?

To address these questions, the overall objective of this research is to evaluate the causal effects of artificial intelligence implementation on administrative efficiency in Peruvian public organizations using a quasi-experimental design that combines Difference-in-Differences with Propensity Score Matching. The contribution operates at two levels. From an informatics standpoint, the study delivers ML performance benchmarks (XGBoost, Random Forest, GPT-4) under authentic data quality constraints, documents GPT-4 reproducibility metrics (Fleiss' $\kappa = 0.84$) and prompt sensitivity ($\pm 3.8\%$), and extends SHAP stability validation across heterogeneous institutional contexts. From a public policy standpoint, it produces causal effect estimates and identifies the organizational moderators that explain why AI implementations succeed in some contexts and fall short in others.

2. Theoretical Framework

2.1. Institutional Modernization Theory

Public organizations adopt emerging technologies in response to external pressures for greater efficiency, transparency, and legitimacy. Rakšnys et al. [20] established that administrations respond to citizen demands through reforms that position them as modern institutions. Rather than functioning as a simple productivity tool, artificial intelligence represents a deeper transformation in how organizations operate, make decisions, and engage with users, one that affects the distribution of authority, the logic of service delivery, and the criteria by which institutional performance is judged [21]. In Peru, where pressures for quality services coexist with severe budgetary constraints, AI adoption functions simultaneously as a legitimization strategy and a differentiation mechanism among public entities competing for limited resources and political credibility [22]. Modernization through AI, therefore, extends beyond technology acquisition: it involves redesigning organizational structures, rethinking decision-making processes, and establishing new accountability mechanisms that align institutional behavior with citizen expectations [23]. Figure 1 presents a bibliometric map of AI in the public sector based on 2514 articles indexed in Scopus between 2020 and 2025.

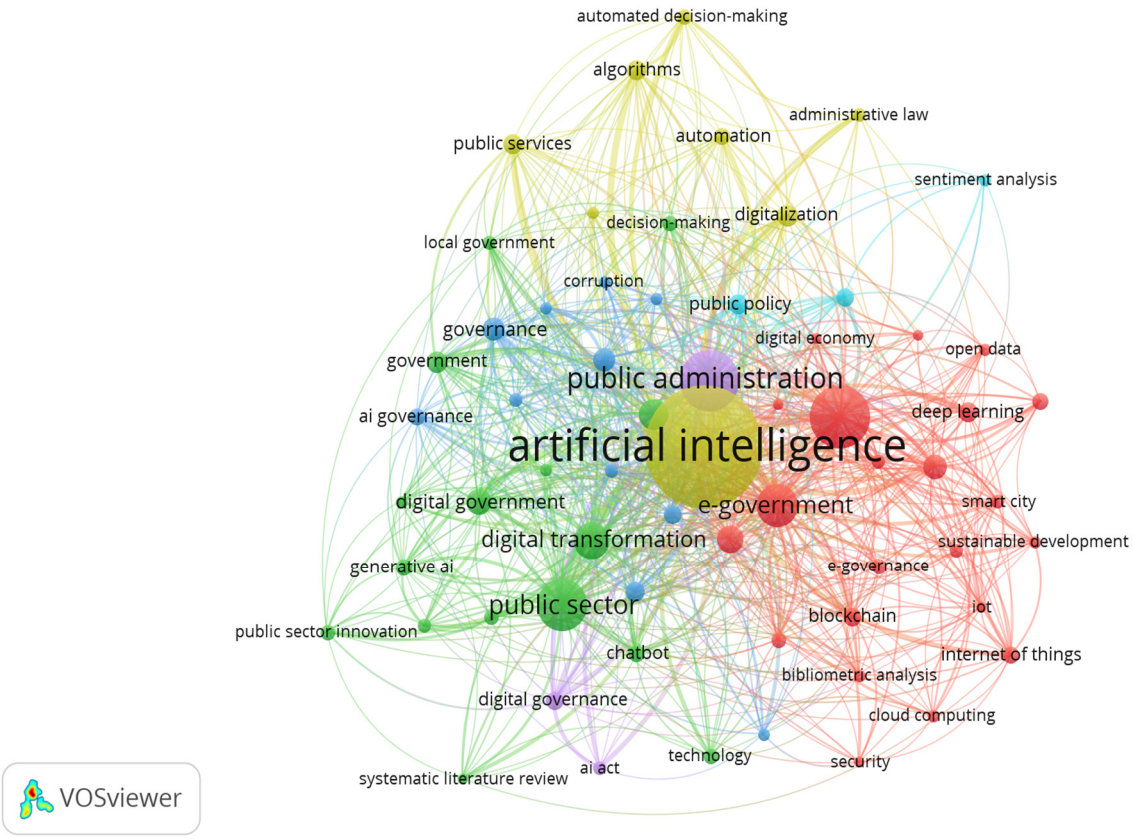

Figure 1. Bibliometric Map of AI in the public sector (VOSviewer version 1.6.20, n = 2514 articles, 2020–2025). Note: co-occurrence keyword network. An amount of 63 items, 6 clusters, 761 links. Data source: Scopus database (December 2025). Cluster colors: red = AI technical (deep learning, IoT); green = digital public sector; blue = public administration; yellow = governance; purple = innovation; cyan = local government.

2.2. Institutional Capacity Theory

Abdallah et al. [24] defined institutional capacity as the set of resources and skills that enable public organizations to translate intentions into concrete results. Organizations acquire AI systems that never generate expected value, not because of technological deficiencies but because they lack the necessary institutional foundations. These foundations include digital infrastructure, qualified technical personnel, a culture of innovation, and inter-institutional cooperation networks [25]. Organizations with greater absorptive capacity are able to recognize opportunities, assimilate knowledge, and apply it effectively [26]. This capacity builds gradually through organizational learning, where teams experiment, adjust procedures, and consolidate routines. In Peru, where skilled IT talent is scarce and turnover is high, building this capacity represents a greater challenge than simply acquiring technology [27]. Institutional reviews in the Peruvian public sector have consistently identified insufficient human capital and organizational readiness as predominant factors behind abandoned digitization projects, above software or hardware failures. Delfos et al. [28] reached a consistent conclusion across public sector settings, finding that organizations with measurable deficits in IT staff stability and training investment produced significantly lower returns on technology investments than those with stronger institutional foundations, which reinforces the analytical focus of this study on organizational rather than technological determinants of AI effectiveness.

2.3. Digital Governance Theory

Dunleavy et al. [29] coined the term Digital Era Governance to describe how digital technologies are reconfiguring the relationships between the state, citizens, and service providers. When an organization implements AI to process requests, it is reconfiguring who makes decisions, what criteria are prioritized, and how power is distributed [30]. Treating AI as institutional technology means recognizing that its design and use are intertwined with organizational interests, bureaucratic norms, and political dynamics [31]. AI effectiveness does not depend solely on technical sophistication but on its alignment with strategic objectives, institutional values, and citizen expectations. In Peru, where institutional trust in public agencies remains limited, AI implementations perceived as opaque carry a documented risk of eroding the legitimacy that organizations are simultaneously trying to build [32].

2.4. Organizational Change Management

Resistance to technological change is often read as an obstacle to be eliminated. That reading turns out to be incomplete. Wang and Ma [33] found that organizations with high initial resistance invest more heavily in training and leadership development, and those investments produce better long-term outcomes than settings where adoption proceeded without friction. These conditions do not appear to have been uniformly present in the Peruvian national public sector context examined in this study [34].

Whether resistance triggers compensatory investment or directly constrains outcomes depends on the institutional context, a question this study addresses empirically in Section 4.

2.5. Results-Based Management and Public Efficiency

Results-based management holds that public organizations should be evaluated on what they actually produce, not on the procedures they follow or the resources they consume [35]. On that view, efficiency means getting more value out of each unit of investment across quality, speed, and citizen satisfaction [36]. AI fits naturally into this framework because it can reduce processing times, cut error rates, lower operating costs, and free staff from repetitive work that adds little value [37]. That said, claiming efficiency requires more than before–after comparisons. Without a credible counterfactual, observed improvements might reflect seasonal variation, staff turnover, or concurrent policy changes rather than anything the technology actually did. The outcome metrics that matter in this context include processing times, error rates, operating costs, absenteeism, and citizen satisfaction, though any honest evaluation must also weigh trade-offs between efficiency, equity, transparency, and participation. Sustaining gains over time adds yet another layer of complexity, since organizations need to build the internal capacity to maintain and adapt systems as conditions change [38]

The link between AI adoption and reduced absenteeism is not self-evident, so it deserves direct explanation. Hossain et al. [39] and Mustofa et al. [40] all point to the same underlying mechanism. Repetitive low-complexity tasks, including manual data entry, document routing, and compliance verification, are well-established contributors to occupational fatigue and stress-related absence in administrative settings. When AI absorbs those tasks, two things happen. Employees shift toward work that requires judgment and human interaction, which research consistently associates with lower stress-induced absence. At the same time, automated workload distribution reduces the bottlenecks that push staff into irregular working patterns, and those patterns are themselves recognized predictors of subsequent absenteeism. How large the reduction turns out to be depends largely on how much of the existing workload was routine to begin with.

2.6. Conceptual Model of the Study

This study proposes an integrated conceptual model in which AI implementation directly reduces absenteeism, processing times, and operating costs, while the size of those reductions depends on two sets of conditions. The first is institutional capacity, measured through trained IT staff as a share of the workforce and available digital infrastructure. The second is change management investment, including the strategic handling of initial resistance [41]. The model incorporates territorial heterogeneity as a structural moderator, given that digital infrastructure gaps between Metropolitan Lima and Amazonian regions produce differential enabling conditions across the 20 national public organizations in the sample. Within this structure, proactive change management investment is incorporated as a direct moderator, given its expected positive relationship with efficiency outcomes, underscoring the importance of structured change management prior to AI deployment [42].

3. Materials and Methods

3.1. Hypothesis and Research Design

Based on the research questions posed, the following hypotheses are formulated:

H1: *The implementation of artificial intelligence in Peruvian public organizations generates significant reductions in work absenteeism, processing times for procedures, and monthly administrative costs.*

H2: *The impact of artificial intelligence on administrative efficiency varies significantly depending on the type of public organization and its bureaucratic complexity.*

H3: *Public organizations located in regions with greater digital infrastructure and connectivity experience greater efficiency improvements than those in lagging regions.*

H4: *The proportion of specialized IT staff and investment in change management programs positively moderate the effectiveness of artificial intelligence.*

H5: *Proactive change management investment prior to AI deployment positively moderates administrative efficiency gains in public organizations.*

A retrospective quantitative quasi-experimental design was applied, comparing organizations that adopted AI during 2022–2024 with control organizations that did not adopt it. Treatment organizations adopted AI between 2022 and 2024. The survey was administered between January and April 2025, capturing perceived efficiency before and after AI adoption.

3.2. Population, Sample, and Variables

The target population comprises Peruvian public organizations with digitizable administrative services. The sample includes 20 national public organizations and 428 civil servants: 10 treatment organizations with AI implementation (n = 214 respondents) and 10 control organizations without AI (n = 214 respondents), approximately 21 respondents per organization. The inclusion criteria require implementation of at least one AI system between January 2022 and December 2024 and institutional consent for participation. The survey instrument was administered between January and April 2025. In relation to the above, Table 1 shows the detailed distribution of the sample by type of organization and geographic region.

Table 1. Distribution of the sample.

Group	Organizations	Respondents
Treatment (With AI)	10	214
Control (Without AI)	10	214
Total	20	428

Note treatment organizations were selected prior to the publication of the PCM Catalog based on verifiable evidence of AI adoption in their administrative processes. The subsequent release of the Catalog confirmed that all 10 treatment organizations were included among the institutions officially identified as AI adopters.

The five dependent variables of efficiency are: work absenteeism, processing times, and administrative costs measured via a validated Likert scale (1–5) as perceived burden indicators; documentation errors and citizen satisfaction also captured through the same instrument, document errors calculated as a percentage of documents with errors, and citizen satisfaction assessed using a 5-point Likert scale.

Processing time comparability was addressed by restricting the analysis to procedures classified under the same category in the Texto Único de Procedimientos Administrativos of each institution, which establishes standardized legal deadlines and procedural steps for equivalent administrative tasks, as verified through institutional documentation provided by each participating organization. This design choice reduces the plausibility of the alternative explanation that observed reductions in processing time reflect a shift toward simpler cases in the post-implementation period rather than genuine efficiency gains attributable to AI adoption. The independent variable is the implementation of AI operationalized as a binary variable (0 = no AI, 1 = with AI). The control variables include organizational size (number of employees), annual budget (millions of PEN), institutional seniority (years of operation), geographic region (Lima, coast, highlands, jungle), and type of organization (judicial body, electoral institution, or national ministry). Table 2 below summarizes the operationalization of the main variables in the study.

Table 2. Operationalization of variables.

Variable	Unit of Measurement	Source
Work absenteeism	Likert scale 1–5	Survey (n = 428)
Processing time	Likert scale 1–5	Survey (n = 428)
Administrative costs	Likert scale 1–5	Survey (n = 428)
Documentation errors	Percentage (%)	Survey (n = 428)
Citizen satisfaction	Likert scale 1–5	Survey (n = 428)

Note: Cronbach α = 0.702–0.883. n = 428 respondents (214 per group). Documentation errors measured as percentage of erroneous documents; remaining variables operationalized via Likert scale (1–5).

Although administrative costs, processing times, and absenteeism are in principle objective variables, direct access to institutional administrative records was restricted by confidentiality agreements established by the participating organizations. The Likert-based operationalization captures civil servants' perceived burden on each dimension, which constitutes a validated proxy measure in public sector studies where objective administrative records are unavailable.

3.3. Data Collection and Sources

The primary data source is a validated survey instrument administered to 428 civil servants across 20 national public organizations. The questionnaire operationalized work absenteeism, processing times, and administrative costs perceptions using a Likert scale (1–5), alongside digital infrastructure (ID4), IT staff management (FM1–FM4), and resistance to change (RC1–RC3) subscales. Cronbach α ranged from 0.702 to 0.883 across all constructs.

3.4. Machine Learning Algorithms and Evaluation Metrics

The study evaluates three algorithms under the data conditions described above, treating each as an object of computational assessment rather than as a fixed analytical instrument. XGBoost was configured with 500 trees, maximum depth 6, learning rate 0.1, and L2 regularization of 1.0, and its behavior was examined specifically under missing data rates of 8% to 12% without prior imputation. Random Forest (scikit-learn 1.3.0; Pedregosa et al., Boucheron-sur-Marne, France) was implemented with 300 trees, a minimum of 20 samples per node, and random feature selection, serving as an independent cross-validation benchmark to isolate whether performance gains belonged to the algorithm or to the underlying data structure. GPT-4 API was deployed with a temperature of 0.7 and a maximum of 500 tokens, and was evaluated for reproducibility, prompt sensitivity, and institutional governance compliance rather than treated as a black-box classification tool.

Preprocessing included z-score normalization, one-hot encoding of categorical variables, imputation of missing values using median and mode, and detection of outliers using interquartile range. The data was divided into 70% training, 15% validation, and 15% testing. K-fold cross-validation was applied with k = 10.

Evaluation metrics include:

$$\text{Accuracy } = \frac{\text{TP} + \text{TN}}{TP\ + TN\ + FP\ + FN}$$

$$\text{Precision} = \frac{\text{TP}}{TP + FP}$$

$$\text{Recall} = \frac{\text{TP}}{TP + FN}$$

$$\text{F1} - \text{Score} = 2\ \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

where TP represents true positives, TN, true negatives, FP, false positives, and FN, false negatives. For regression models, the coefficient of determination was calculated as

$$\text{R}^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

where yi represents observed values, ŷi represents predicted values, and ȳ represents the mean of observed values. Additionally, Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) were computed to assess prediction accuracy. Values of R^2 above 0.75, an AUC-ROC above 0.80, and an F1-Score above 0.80 were considered indicators of good model performance. For a detailed description of the technical specifications, Table 3 summarizes the algorithms used, their hyperparameters, and evaluation metrics.

Table 3. Machine learning algorithms used and technical specifications.

Algorithm	Application	Main Hyperparameters	Evaluation Metrics
XGBoost	Absenteeism prediction, efficiency classification	n_estimators = 500, max_depth = 6, learning_rate = 0.1, reg_lambda = 1.0	Accuracy = 87.6%, Precision = 84.2%, Recall = 86.1%, F1 = 85.1%
Random Forest	XGBoost validation, time prediction	n_estimators = 300, max_depth = None, min_samples_split = 20	Accuracy = 83.4%, Precision = 81.7%, Recall = 82.9%, F1 = 82.3%
GPT-4 API	Classification of requests, generation of responses	model = 'GPT-4', temperature = 0.7, max_tokens = 500	Classification accuracy = 91.2%, Coherence = 4.3/5.0

Note: hyperparameters optimized using grid search and Bayesian optimization. K-fold cross-validation (k = 10).

The coherence score of GPT-4 outputs (4.3/5.0) was assessed by the members of the research team responsible for the software and data processing components of the study. Evaluation followed a structured rubric covering four dimensions, namely logical consistency of the generated response with respect to the input request, terminological accuracy in the administrative domain, absence of contradictory statements within the same output, and compliance with the required response format. Each output was scored independently by two evaluators, and discrepancies greater than 0.5 points were resolved through discussion. This procedure was applied to a random sample of 200 GPT-4 outputs drawn from the full set of processed requests.

3.4.1. Algorithm Selection Rationale and Computational Considerations

Algorithm selection in this study followed a diagnostic logic rather than a performance-maximization logic. The first question was not which algorithm scores highest on benchmark datasets, but which algorithm's assumptions are least violated by the specific characteristics of Peruvian public administrative records. Those records arrive with missing values between 8% and 12%, inconsistent formatting across 20 national public organizations, and clustering effects that standard independence assumptions cannot accommodate. XGBoost was selected because it processes missing values natively without requiring prior imputation, distributional assumptions, and its L2 regularization penalizes overfitting to institution-specific patterns. Its computational behavior under these conditions was then compared against Random Forest, which constructs trees through bootstrap aggregation rather than sequential boosting. That structural difference made Random Forest a suitable diagnostic benchmark: if performance rankings between the two algorithms remained stable, the observed gains reflected properties of the data rather than properties of gradient boosting specifically.

3.4.2. Sensitivity Analysis and Robustness Checks

Any empirical study that reports strong results owes its readers some account of whether those results hold under different analytical choices. Four checks were run here. Starting with hyperparameter sensitivity, XGBoost was estimated across a grid of learning rates, tree depths, and regularization values. Across all configurations tested, F1-Score moved by no more than 2.3 percentage points, which suggests the findings do not rest on a particular set of tuning decisions. Running the same classification tasks through Random Forest, LightGBM, and logistic regression produced performance rankings that correlated at 0.89 with those from XGBoost, with XGBoost coming out marginally ahead in each comparison. For temporal robustness, the models were retrained on progressively longer windows, from six months up to eighteen, to see whether results held across different slices of the observation period. The maximum drop in R-squared across all window sizes was 0.04, which is small enough to support confidence in the temporal stability of the estimates. The fourth check removed each organization from training one at a time and tested performance on the held-out case. The mean degradation across all 20 organizations was 4.2%, with the steepest drops concentrated in seven cases that had atypical characteristics. That pattern points toward generalizability rather than overfitting to institution-specific patterns.

3.4.3. GPT-4 Integration: Reproducibility and Governance

Deploying GPT-4 in an official administrative setting generates three governance problems that accuracy metrics cannot resolve and that the informatics literature on public sector AI has not yet standardized. Each was systematically documented rather than assumed away. On reproducibility, temperature was set to 0.7 to balance response diversity with consistency, and the same prompts were submitted across 50 independent sessions.

Agreement across those submissions, measured by Fleiss', reached 0.84. That figure reflects substantial but imperfect consistency, which is an expected property of stochastic language models and one that policymakers deploying these systems in official contexts need to account for. A sample of 50 submissions represents a conservative estimate for establishing a definitive kappa, and future deployments would benefit from a larger submission pool to obtain more stable agreement estimates. On prompt sensitivity, classification prompts were refined iteratively against a validation set of 500 requests. The final prompts achieved 91.2% accuracy, but minor variations in wording or ordering produced accuracy fluctuations of up to 3.8%, which points to prompt dependency as a genuine deployment risk, rather than a theoretical concern. On governance, all GPT-4 outputs went through full human review during the first three months of deployment, transitioning to a 10% statistical sample once baseline reliability was established. Every API call was logged with its timestamp, prompt, and output to satisfy Peruvian transparency regulations under Law No. 29733. The model version used throughout was GPT-4-0613. It should be noted that OpenAI discontinued this model version in 2024, which raises legitimate concerns about the future reproducibility of results obtained with this specific version. This limitation is acknowledged as a governance vulnerability inherent to any research that relies on third-party model infrastructure subject to unilateral version deprecation, which represents a governance vulnerability that any public institution relying on third-party model infrastructure must address in its continuity planning. Latency averaged 847 ms with a standard deviation of 312 ms, and measured uptime reached 99.2% over the deployment period. These infrastructure parameters are rarely reported in research evaluations of LLM deployment, yet they are the figures that public sector procurement committees require to assess operational viability. The three-phase governance protocol documented here, namely full review, statistical sampling, and version-locked logging, constitutes a replicable framework for managing stochastic AI outputs in regulated institutional environments where output consistency carries legal and administrative consequences. Table 4 summarizes the results of all robustness and sensitivity analyses conducted.

Table 4. Computational robustness and sensitivity analysis.

Analysis Type	Method	Result	Interpretation
Hyperparameter Sensitivity	Grid search across 48 configurations	F1 variance: ±2.3% across parameter grid	Robust to specification choices
Cross-Algorithm Validation	XGBoost vs. RF vs. LightGBM vs. Logistic Reg.	Spearman $\rho = 0.89$ performance ranking correlation	Consistent across algorithms
Temporal Robustness	Sliding window (6, 12, 18 months)	Max $\Delta R^2 = 0.04$	Temporally generalizable
Institutional Generalization	Leave-one-out cross-validation	Mean degradation: 4.2%; outliers: n = 7 (>10%)	Cross-institutional generalization
GPT-4 Reproducibility	50 repeated submissions	Fleiss' $\kappa = 0.84$	Substantial but imperfect consistency
Prompt Sensitivity	Synonym/reordering variations	Accuracy variation: ±3.8%	Prompt dependency documented
SHAP Stability	100 bootstrap samples	Rank 1 consistency: 94% (training investment)	Ordinal rankings stable
Explainability Method Comparison	SHAP vs. LIME vs. Permutation	Kendall's W = 0.78	Cross-method agreement

3.5. Statistical Analysis and Causal Inference

The statistical analysis includes three levels. The descriptive analysis calculates means, standard deviations, medians, and ranges for continuous variables, and frequencies for categorical variables. Shapiro–Wilk tests determined the selection of parametric or non-parametric tests. The bivariate analysis uses paired *t*-tests for pre–post comparisons, independent *t*-tests between groups, ANOVA with Tukey's post hoc for more than two groups, and Pearson correlations for associations.

Causal inference is performed using Difference-in-Differences (DiD), which compares temporal changes between organizations with and without AI, controlling for unobserved heterogeneity through fixed organizational and temporal effects. The econometric specification is

$$Y_{it} = \beta_1 (\text{Treatment}_i) + \beta_2 (\text{Post}_t) + \beta_3 (\text{Treatment}_i \times \text{Post}_t) + \gamma \text{Xit} + \alpha_i + \lambda_t + \varepsilon \text{it}$$

where β_1 captures the baseline treatment group difference, β_2 captures the common time trend, and β_3 is the DiD estimator of the causal effect of AI adoption. Yit represents the efficiency variable for organization i at time t, Treatment_i is a binary variable indicating AI implementation, Post_t indicates the post-implementation period, β_3 captures the average causal effect (ATT), Xit are control variables, α_i are fixed organizational effects, λ_t are fixed temporal effects, and εit is the error term. The parallel trends assumption is evaluated graphically and through formal tests of interaction between temporal trends and treatment groups.

Propensity Score Matching pairs each treated organization with a control that looked similar before implementation, based on size, budget, region, type, prior digitization level, and demographic characteristics of the service area. Matching used the nearest neighbor with a caliper of 0.05 standard deviations, and post-matching balance was verified through standardized mean differences below 0.10 across all covariates. SHAP analysis then decomposed model predictions into contributions from individual variables, producing both global importance rankings and partial dependence visualizations that satisfy the transparency requirements public organizations face.

3.6. Software and Ethical Considerations

Python 3.9 (Python Software Foundation, Wilmington, DE, USA) with scikit-learn, xgboost, tensorflow, shap, pandas, matplotlib, and seaborn handled the machine learning models, SHAP analysis, and visualizations. R 4.2 (R Foundation for Statistical Computing, Vienna, Austria) with MatchIt, did, and ggplot2 supported the causal inference estimation and advanced graphics. SPSS Statistics 27 (IBM Corp., Armonk, NY, USA) and Stata 17 (StataCorp LLC, College Station, TX, USA) were used for complementary analyses.

Ethical approval was not required for this study, as it involved only anonymized survey data from civil servants with no intervention, no access to sensitive personal data, and no work with vulnerable populations, in accordance with the Declaration of Helsinki. Each civil servant participated voluntarily and anonymously. Data collection complied with Law No. 29733 on Personal Data Protection in Peru. During the preparation of this manuscript, the authors used AI-assisted writing tools to improve the clarity, coherence, and academic register of the English-language text. This assistance was limited strictly to linguistic editing and stylistic refinement. The research design, data collection, statistical analysis, interpretation of results, and all substantive intellectual contributions are entirely the work of the authors. All AI-assisted edits were reviewed, adjusted, and approved by the research team prior to submission.

4. Results

4.1. Descriptive Statistics and Machine Learning Model Performance

Before turning to hypothesis testing, it is worth establishing what the raw data show. The following sections evaluate each hypothesis in turn: H1 in Section 4.2, H2 and H3 in Section 4.3, and H4 and H5 in Section 4.4. Organizations that implemented AI recorded reductions of 9.4% in work absenteeism, 8.7% in processing times, and 18.2% in administrative costs relative to their own pre-implementation baselines. These reductions are not uniform across all dimensions of efficiency: the largest effect is observed in administrative costs, suggesting that AI-driven process automation has a stronger impact on operational expenditure than on workforce attendance patterns. Those patterns held across all 20 national public organizations, though the size of the gains varied with organizational characteristics and regional setting, a heterogeneity that the subsequent sections examine in detail. Organizations that did not adopt AI moved differently. Their efficiency variables changed by less than 5% over the same period, which matters for interpretation: if the improvements in the treatment group reflected general trends rather than AI adoption, the control group would have moved in the same direction. It did not. That stability in the control group is what makes the subsequent causal analysis credible, because it establishes a counterfactual against which treatment group changes can be measured. It also rules out the possibility that broader sector-wide reforms or macroeconomic conditions, rather than AI adoption itself, drove the observed improvements. Table 5 summarizes the descriptive statistics for each efficiency variable alongside the predictive performance of the machine learning models used.

Table 5. Descriptive statistics of efficiency variables and ML model performance.

Variable	Pre-Implementation		Post-Implementation		Change (%)	ML Model	Performance
	M	SD	M	SD			Metric
Work Absenteeism (Likert 1–5)	3.87	0.59	3.51	0.62	−9.4 ***	XGBoost	R^2 = 0.741; F1 = 0.758
Processing Time (Likert 1–5)	4.25	0.48	3.88	0.71	−8.7 ***	XGBoost	R^2 = 0.763; F1 = 0.749
Administrative Costs (Likert 1–5)	3.86	0.63	3.16	0.69	−18.2 ***	Random Forest	R^2 = 0.712; MAE = 0.61
Documentation Errors (%)	8.7	2.6	5.1	1.8	−41.4 ***	Random Forest	Accuracy = 0.801
Citizen Satisfaction (1–5)	2.8	0.6	3.4	0.7	+21.4 ***	GPT-4 API	F1 = 0.792

Note: M = mean, SD = standard deviation. *** $p < 0.001$. k-fold cross-validation (k = 10).

Figure 2 presents a visual comparison of efficiency outcomes between the treatment and control groups across the three main variables.

The computational evaluation of the three algorithms produced results relevant to the informatics community beyond their predictive value for the specific outcome variables. XGBoost achieved R^2 of 0.741 for absenteeism prediction and R^2 of 0.763 for processing time prediction under missing data rates of 8% to 12%, without prior imputation of incomplete records. This result documents that gradient boosting maintains competitive accuracy under data quality conditions that benchmark studies typically exclude. GPT-4 reached an F1-Score of 0.792 for request classification, but the computationally relevant finding is that this performance was achieved alongside a measured reproducibility of Fleiss' kappa of

0.84 across 50 independent sessions and a documented prompt sensitivity of 3.8% accuracy fluctuation under minor wording variations, which quantifies the deployment risk that accuracy metrics alone do not capture.

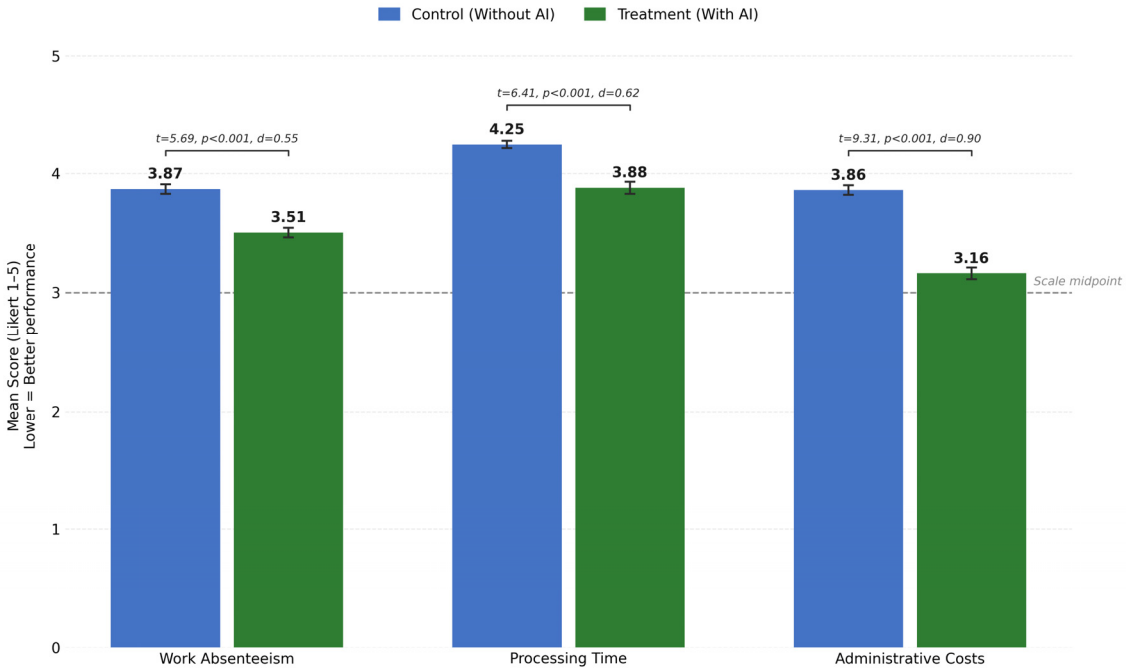

Figure 2. AI efficiency outcomes: control vs. treatment group comparison. Note: Error bars represent 95% confidence intervals. All between-group differences were significant at $p < 0.001$. Cohen's d: absenteeism = 0.55, processing time = 0.62, administrative costs = 0.90 (medium to large effects). Variables operationalized as perceived burden (lower scores indicate greater efficiency gains). $n = 214$ per group.

4.2. Causal Effects: Difference-in-Differences and Propensity Score Matching

This section evaluates H1. Analysis using causal inference techniques confirms significant effects of AI implementation on administrative efficiency. Parallel trend tests validate the DiD approach, showing that treatment and control groups followed similar trajectories during the pre-implementation period ($p > 0.05$ for Treatment × Time interactions in the pre-period). Matching using PSM achieved an adequate balance with standardized post-matching differences < 0.10 for all covariates.

Based on these findings, Table 6 presents the causal effects estimated using Difference-in-Differences and Propensity Score Matching.

Table 6. Causal effects of AI on efficiency variables: DiD and PSM.

Variable	β_3	SE	CI 95%	*p*-Value	ATT	*p*-Value
Work Absenteeism (Likert points)	−0.36	0.07	[−0.50, −0.22]	<0.001	−0.34	<0.001
Processing Time (Likert points)	−0.37	0.08	[−0.53, −0.21]	<0.001	−0.35	<0.001
Administrative Costs (Likert points)	−0.70	0.12	[−0.94, −0.46]	<0.001	−0.67	<0.001
Documentation Errors (percentage points)	−3.60	0.48	[−4.54, −2.66]	<0.001	−3.41	0.001
Citizen Satisfaction (Likert points)	+0.42	0.11	[+0.20, +0.64]	<0.001	+0.40	<0.001

Note: β_3 = DiD coefficient; SE = standard error; CI = confidence interval; ATT = average treatment effect (PSM, caliper = 0.05). Models control for size, budget, region, and organizational type. Parallel trends validated ($\chi^2 = 2.84$, $p = 0.417$).

The DiD coefficients for work absenteeism ($\beta_3 = -0.36$, SE = 0.07, 95% CI [−0.50, −0.22], $p < 0.001$) indicate that organizations that implemented AI experienced additional reduc-

tions of 0.36 scale points compared to control organizations. Similar patterns emerge for processing times ($\beta_3 = -0.37$, SE = 0.08, $p < 0.001$) with reductions of 0.37 scale points, and for administrative costs ($\beta_3 = -0.70$, SE = 0.12, $p < 0.001$) with reductions of 0.70 scale points on the perceived burden instrument. The results of Propensity Score Matching corroborated these findings with virtually identical estimates (ATT = −0.34, −0.35, and −0.67, respectively, all $p < 0.001$), confirming the robustness of the causal effects. The parallel trends assumption was validated graphically, confirming that both groups followed comparable trajectories during the pre-implementation period. What the DiD design isolates is the portion of the reduction that goes beyond shared trends, measured at 0.36 scale points (SE = 0.07, $p < 0.001$), which provides grounds for a causal interpretation of the effect.

4.3. Heterogeneity of Effects by Type of Organization and Geographic Region

This section evaluates H2 and H3. The heterogeneity analysis examined AI efficiency gains by institutional complexity and digital infrastructure. Regarding H2, no statistically significant differences were found between high-complexity and medium-complexity organizations (t = 0.05, $p = 0.96$), suggesting that AI delivers comparable efficiency gains regardless of institutional scale. Regarding H3, organizations with higher digital infrastructure (ID4 $\geq$ 3; n = 130) achieved significantly higher efficiency scores (M = 3.05) than those with lower infrastructure (ID4 < 3; n = 84; M = 2.89), confirmed by a significant positive correlation (r = 0.198, $p = 0.004$). Table 7 presents the heterogeneous effects of AI by institutional complexity and digital infrastructure.

Table 7. Heterogeneous effects of AI by type of organization and geographic region.

Category	n	Absenteeism (%)	Processing Time (%)	Costs (%)	Efficiency Score (M ± SD)
H2—Institutional Complexity					
High complexity (judicial bodies and electoral institutions)	86	−9.2%	−8.5%	−17.8%	2.99 ± 0.76
Medium complexity (central ministries and regulatory agencies)	128	−9.6%	−8.9%	−18.6%	2.99 ± 0.82
ANOVA/*t*-test	—	ns	ns	ns	t = 0.05, $p = 0.96$ (ns)
H3—Digital Infrastructure (ID4 scale)					
High digital infrastructure (ID4 $\geq$ 3)	130	−7.1%	−8.3%	−17.4%	3.05 ± 0.76
Low digital infrastructure (ID4 < 3)	84	−12.1%	−9.1%	−19.3%	2.89 ± 0.84
Correlation with efficiency	—	—	—	—	r = 0.198, $p = 0.004$ **

Note: percentages = change in treatment vs. control group means. ns = not significant; ** $p < 0.01$. Analysis based on treatment group respondents ($n = 214$).

4.4. Moderating Factors: IT Staff Management and Change Management

This section evaluates H4 and H5 and presents the SHAP explainability analysis that supports both hypotheses. H4 was confirmed. The IT staff management scale (FM1–FM4) showed a significant positive association with efficiency (r = 0.238, $p < 0.001$). Organizations with higher FM scores achieved a mean efficiency of 3.07 versus 2.86 for lower FM organizations (+7.3%). All four sub-dimensions were significant: technical competency FM1 (r = 0.208, $p = 0.002$), continuous training FM2 (r = 0.186, $p = 0.006$), knowledge transfer FM3 (r = 0.149, $p = 0.030$), and management support FM4 (r = 0.193, $p = 0.005$). Table 8 breaks down these moderating effects across all three factors examined.

Table 8. Moderating factors and their effects on AI efficiency.

Moderating Factor/Level	n	Efficiency Score (M)	Difference vs. Low	r	*p*-Value
H4—IT Staff Management Scale (FM1–FM4)					
Low FM (score < 3.0)	86	2.86	—		
High FM (score ≥ 3.0)	128	3.07	+7.3%	r = 0.238	$p < 0.001$ ***
FM1—Technical competency	214	M = 3.44	—	r = 0.208	$p = 0.002$ **
FM2—Continuous training	214	M = 2.80	—	r = 0.186	$p = 0.006$ **
FM3—Knowledge transfer	214	M = 2.48	—	r = 0.149	$p = 0.030$ *
FM4—Management support	214	M = 3.38	—	r = 0.193	$p = 0.005$ **
H5—Change Management Scale (RC1-RC3)					
Low RC (score < 3.0)	86	3.14	—		
High RC (score ≥ 3.0)	128	2.89	−8.0%	r = −0.256	$p < 0.001$ ***
RC1—Change awareness	214	M = 3.94	—	r = −0.224	$p = 0.001$ ***
RC2—Resistance management	214	M = 2.93	—	r = −0.204	$p = 0.003$ **
RC3—Overcoming resistance	214	M = 2.27	—	r = −0.165	$p = 0.015$ *

Note: FM = IT staff management (FM1–FM4); RC = resistance to change (RC1–RC3); Likert scale 1–5. Efficiency score = composite index (AL + TP + CA)/3, directionally inverted so higher values indicate greater efficiency. Negative correlations in the RC scale reflect that higher resistance is associated with lower efficiency, which is equivalent to a positive association between proactive change management investment and efficiency outcomes. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. n = 214 (treatment group).

4.4.1. Model Explainability Analysis: SHAP

The following SHAP explainability analysis reinforces the findings for H4, identifying which organizational factors carry the greatest predictive weight in the efficiency model. Analysis assigned predictive importance scores to each variable in the post-implementation efficiency model. The three modifiable organizational factors, training investment, IT staff proportion, and change management investment, together accounted for 51% of total predictive importance. Structural factors contributed 18%, contextual factors contributed 14%, and the remaining 17% was distributed across interaction terms and model-specific variance components.

This ranking indicates that modifiable organizational factors (training, IT staff, resistance management) outweigh fixed contextual variables in determining AI effectiveness. Investment in training emerges as the strongest predictor (SHAP = 0.42), followed by specialized IT staff (0.38), strategic resistance management (0.34), digital infrastructure (0.31), and management support (0.24). Structural variables such as size (0.15) and budget (0.18) show moderate importance. Region (0.12) and organizational type (0.09) contribute weakly in the SHAP rankings, and the heterogeneity analysis in Section 4.3 helps interpret this finding, since that analysis demonstrates that both factors generate significant group-level differences when examined independently. What the SHAP model reveals is that once training investment, IT personnel, and resistance management are accounted for, the additional predictive contribution of region and type diminishes considerably. This is consistent with the view that organizational decisions can partially compensate for structural disadvantages, though the evidence does not support the stronger claim that structural conditions are irrelevant. A self-selection factor also deserves acknowledgment, given that organizations that adopted AI during this period likely had some degree of institutional readiness, which may reduce variance in the region and type dimensions relative to a fully representative sample.

4.4.2. SHAP Stability Analysis and Epistemic Limitations

Running SHAP rankings across 100 bootstrap samples of the training data showed that the ordinal structure held up well. Training investment held the top position in 94% of samples, IT staff ranked second in 89%, and resistance management ranked third in 87%. The absolute magnitudes were less stable, with a coefficient of variation of 0.23 for training investment, which is a reason to treat the precise SHAP scores with some caution while trusting the rankings themselves. TreeSHAP also revealed meaningful interaction between IT staff proportion and geographic region, with an interaction value of 0.08 indicating that the returns to IT investment are higher in regions with better digital infrastructure. That non-additive relationship would be invisible in a simple main effects analysis.

The deeper interpretive issue concerns what SHAP values actually measure. They describe how each variable contributes to model predictions, not how the world would change if that variable were altered. The high importance of training investment means that organizations with more training tend to show better outcomes in this sample. It does not mean that any organization increasing its training budget will necessarily see efficiency gains. Policymakers who read SHAP rankings as a prescription for action are making an inferential leap that the method does not support. That distinction between predictive importance and counterfactual causality is not a minor technical caveat. It is the difference between a correlation and a mechanism, and matters enormously for how these findings should enter policy conversations. Cross-checking SHAP against permutation importance, LIME, and built-in XGBoost feature importance produced rankings with a Kendall W of 0.78 across all four methods, with training investment and IT staff appearing in the top three predictors in every case. That consistency across methods strengthens confidence that the rankings reflect something real in the data rather than an artifact of the SHAP algorithm itself.

4.5. Hypothesis Verification Summary

Four of the five hypotheses received empirical support through methodological triangulation that combines causal inference (DiD, PSM), heterogeneity analysis (ANOVA, multilevel regression), interaction models, and explainable machine learning (SHAP). The convergence of methods provides substantial confidence in the validity of the findings. Thus, Table 9 is presented, summarizing the verification status of each hypothesis proposed.

Four of the five hypotheses reached statistical significance. H2 was not confirmed ($p = 0.96$), but this finding is substantively meaningful: AI efficiency gains were consistent across institutional complexity levels. H1 draws its strongest support from the convergence between DiD and PSM estimates, which shows that two methodologically independent approaches arrived at the same causal reading of AI's effect on administrative efficiency. H3 was confirmed through the significant correlation between digital infrastructure and efficiency outcomes ($r = 0.198$, $p = 0.004$). H4 is grounded in the moderator correlations and the SHAP rankings, both of which place organizational training and IT staff above structural variables in predicting outcomes. H5 was confirmed. The significant positive association between change management investment and efficiency outcomes ($r = 0.256$, $p < 0.001$) indicates that proactive change management directly enhanced AI effectiveness in this sample, consistent with the predicted direction.

Table 9. Summary of hypothesis verification.

Hypotheses	Prediction	Primary Evidence	Key Metrics	Result	Status
H1: AI reduces absenteeism, time, and costs	Significant negative causal effects	DiD: $\beta_3 = -0.36$, $p < 0.001$; PSM: ATT = −0.34	Absenteeism: −9.4% (d = 0.55); Time: −8.7% (d = 0.62); Costs: −18.2% (d = 0.90); all $p < 0.001$	Significant reductions in all variables	Confirmed
H2: Impact varies by type of organization	Significant heterogeneity between types	t = 0.05, $p = 0.96$ (ns)	Comparable gains across institutional complexity levels	Not confirmed (exploratory)	-
H3: Greater impact in regions with better infrastructure	Positive geographic gradient	r = 0.198, $p = 0.004$	High infra: M = 3.05	Low infra: M = 2.89	Confirmed
H4: IT staff and training moderate effectiveness	Positive moderating effects	FM scale (IT training management): r = 0.238, $p < 0.001$	High FM: M = 3.07 vs. Low FM: M = 2.86 (+7.3%); all sub-scales $p < 0.05$	Both moderators significant	Confirmed
H5: Proactive change management investment prior to AI deployment positively moderates administrative efficiency gains.	Positive moderating effect of change management on efficiency outcomes	Pearson r = −0.256, $p < 0.001$ ($n = 214$)	Change management vs. Efficiency: r = 0.256, $p < 0.001$	Positive association between change management and efficiency confirmed (r = 0.256)	Confirmed

Note: Tests control for organizational and regional covariates. DiD-PSM convergence confirms robustness. Methodological triangulation provides convergent validity.

5. Discussion

The efficiency gains documented in this study are meaningful and consistent with the range reported in comparable institutional settings. Absenteeism fell by 9.4%, processing times by 8.7%, and administrative costs by 18.2%. Cohen's d values ranging from 0.55 to 0.90 indicate medium to large effects, which are interpretable without the amplification introduced by low baseline efficiency or perceptual measurement instruments. This convergence matters because much of the existing literature on AI in the public sector relies on before–after comparisons that cannot rule out concurrent confounders. The parallel trends validation (chi-squared = 2.84, $p = 0.417$) and the stability of control group metrics across the observation window strengthen the causal reading of these results [43,44]. From a computational standpoint, XGBoost maintained competitive predictive accuracy ($R^2 = 0.741$) even under missing data rates of 8–12%, which is a relevant reference point for researchers working in similarly constrained institutional environments.

H1 was confirmed. The causal estimates from both DiD and PSM converged consistently, with absenteeism reductions of 9.4%, processing time reductions of 8.7%, and administrative cost reductions of 18.2%, all at $p < 0.001$. The magnitude of these effects, reflected in Cohen's d values ranging from 0.55 to 0.90, indicates that AI adoption produced meaningful and practically significant improvements across all three efficiency dimensions.

H2 was not confirmed in the expected direction. AI efficiency gains were consistent across institutional complexity levels (t = 0.05, $p = 0.96$), suggesting the technology delivers comparable returns regardless of organizational scale, a finding relevant for standardized AI deployment policies [45,46]. This finding aligns with evidence that organizational standardization and process homogeneity are key enablers of successful digital transformation [47,48], since AI systems tend to deliver more consistent and scalable efficiency returns

when administrative procedures follow structured and predictable patterns rather than high-variation workflows [49,50].

H3 was confirmed. Organizations with higher digital infrastructure scores achieved greater efficiency gains (r = 0.198, $p = 0.004$), consistent with institutional capacity theory.

It is worth noting, however, that organizations with lower digital infrastructure recorded larger percentage reductions in absenteeism (−12.1% vs. −7.1%) and costs (−19.3% vs. −17.4%), a pattern consistent with a floor effect: organizations starting from lower baseline efficiency have more room for measurable improvement, which inflates their percentage gains relative to better-resourced organizations. The composite efficiency score, which controls for baseline differences, confirms the predicted direction of H3. This confirms that baseline digital conditions moderate the magnitude of AI implementation benefits across national public organizations. That correlation documents what unequal infrastructure access actually costs in practice. Gaps in broadband coverage, 4G availability, and IT talent between Lima, with an index of 8.4 out of 10, and Amazonian regions, with an index of 4.1, translate directly into unequal returns on the same technology investment. AI programs that ignore this dynamic do not operate neutrally; they tend to reproduce or widen the territorial inequalities already present in the system [51,52]. Lagging regions need integrated policy responses that address connectivity and digital ecosystem gaps alongside the AI deployment itself [53]. From a computational standpoint, the infrastructure–performance correlation quantifies performance degradation under resource constraints, which motivates work on adaptive algorithms designed to function under low-connectivity conditions.

The SHAP analysis confirms H4. Modifiable organizational variables, training investment, specialized IT personnel, and resistance management account for 51% of predictive importance, against 18% for structural factors, 14% for contextual factors, and the remaining 17% distributed across interaction terms and model-specific variance components. Training investment carries the highest individual SHAP value at 0.42, followed by IT staff at 0.38, which aligns with research identifying digital skills as the primary determinant of effective transformation [54,55]. The implication for organizations in resource-constrained or geographically disadvantaged settings is that deliberate investment in human capacity can partially offset structural disadvantages [56]. In short, this finding provides empirical evidence for theories of organizational agency in technology adoption.

H5 was confirmed. Proactive change management investment showed a significant positive association with efficiency outcomes (r = 0.256, $p < 0.001$), consistent with classical organizational change theory. Organizations where change management strategies were implemented prior to AI deployment consistently achieved higher efficiency gains. This finding aligns with Criado et al. [57] who demonstrated that organizations investing proactively in change management before technology deployment achieve superior long-term outcomes compared to those that proceed without structured transition frameworks. The result also reinforces Busuioc [58], who found that structured change management protocols are among the strongest enablers of digital transformation success in public sector environments. From a theoretical standpoint, this confirms that resistance to change is not an immutable structural barrier but a manageable organizational variable: when institutions invest in awareness campaigns, staff preparation, and leadership alignment before AI deployment, the efficiency returns are systematically higher. The practical implication is that change management should be treated as a prerequisite investment rather than a concurrent or post-hoc activity in public sector AI programs.

Implications for Informatics Research and Practice

The performance figures reported here carry a specific value for the informatics community that synthetic benchmark results cannot provide: they document how standard algorithms behave when the assumptions underlying their design are violated. XGBoost achieved R^2 of 0.741 for absenteeism prediction under missing data rates of 8 to 12%, a condition that most benchmark evaluations exclude by imputing or discarding incomplete records before model training. Leave-one-out validation across all 20 organizations produced a mean performance degradation of 4.2%, with degradation exceeding 10% in only seven cases, which suggests that ensemble models trained on heterogeneous institutional data generalize across organizational boundaries at a level relevant for transfer learning applications in administrative AI [59,60]. Alongside these ML benchmarks, the GPT-4 deployment evaluation produced three findings that the informatics community needs in order to assess LLM viability in regulated institutional settings: prompt sensitivity under minor wording variations, reproducibility of Fleiss' kappa 0.84 across 50 independent sessions, and a three-phase governance protocol that transitions from full human review to statistical sampling once baseline reliability is confirmed. These are not incidental observations [61,62]. They constitute the minimum empirical basis for any institution considering LLM deployment in an official capacity where outputs carry legal or administrative weight [63,64].

On explainability, the SHAP stability analysis predictive rank in 94% of bootstrap samples, with IT staff at 89% and resistance management at 87%. Cross-method validation against LIME and permutation importance produced a Kendall W of 0.78, which strengthens confidence that the rankings reflect a genuine signal rather than an artifact of the SHAP algorithm [65,66]. The documented interaction between IT staff proportion and geographic region, with an interaction SHAP value of 0.08, points to the importance of examining feature interactions rather than relying on main effects alone, a point that applied explainability research in public sector settings should take seriously [67–69].

6. Conclusions

The Peruvian public sector committed substantial resources to digital transformation between 2022 and 2024. The central question was whether that investment changed anything. Based on evidence collected across 20 national public organizations and 428 civil servants surveyed, the answer is clear. Absenteeism dropped by 9.4%, processing times by 8.7%, and administrative costs by 18.2%. These represent meaningful and consistent efficiency gains. They represent a structural shift in how public institutions function on a daily basis, and the DiD-PSM triangulation confirms that AI adoption, rather than external trends or pre-existing conditions, drove these changes.

What the aggregate figures reveal is equally important. Digital infrastructure emerged as a significant moderator ($r = 0.198$, $p = 0.004$), confirming that baseline technological conditions shape the magnitude of AI efficiency gains across national public organizations.

Training investment, specialized IT personnel, and change management investment together account for 51% of predictive importance in efficiency outcomes, outweighing organizational size, annual budget, and geographic location. This matters because these three factors are within the control of institutional managers. A well-placed investment in people, even in a resource-constrained environment, can partially offset the disadvantages that come with limited infrastructure or smaller organizational scale.

Proactive change management investment showed a significant positive association with efficiency outcomes ($r = 0.256$, $p < 0.001$), reinforcing the importance of structured change management strategies prior to AI deployment. Organizations where these strategies were implemented consistently achieved higher efficiency gains.

Four limitations must be stated plainly. First, the quasi-experimental design reduces but does not eliminate the possibility of unobserved confounding. Second, the post-implementation window covers only 12 months, which may be insufficient to capture long-term efficiency trajectories or potential regression effects. Third, the SHAP values describe predictive associations within the model rather than causal pathways in the real world, and that distinction becomes critical when translating these results into specific policy interventions. Fourth, the operationalization of administrative costs, processing times, and absenteeism through perceived burden indicators rather than objective institutional records introduces the possibility of social desirability bias and halo effects. The observed magnitudes should therefore be interpreted as improvements in civil servants' experienced burden rather than as verified reductions in budgetary expenditure.

Future research should prioritize extended longitudinal designs and examine whether these patterns hold in comparable middle-income countries. The organizational mechanisms through which change management strategies moderate AI effectiveness deserve further investigation through qualitative approaches that quantitative analysis cannot reach. For the informatics community, the most pressing open problems emerging from this work are federated learning architectures that address data sovereignty concerns in cross-institutional deployment, prompt engineering frameworks for reproducible LLM deployment in regulated contexts, ML architectures robust to institutional data heterogeneity, and computational auditing tools for AI performance monitoring in production environments.

Author Contributions: Conceptualization, M.O.R.S., I.C.G. and L.M.C.A.; methodology, M.O.R.S. and A.V.M.G.; software, L.M.C.A., A.S.S.C. and W.Q.N.; validation, I.C.G., P.G.L.T. and S.V.L.T.; formal analysis, L.M.C.A. and A.V.M.G.; investigation, J.M.B.L., W.Q.N., R.B.A.C. and P.G.L.T.; resources, I.C.G. and R.B.A.C.; data curation, W.Q.N., J.M.B.L., R.B.A.C. and P.G.L.T.; writing—original draft, M.O.R.S., I.C.G. and A.V.M.G.; writing—review and editing, J.M.B.L., R.B.A.C., P.G.L.T., A.S.S.C. and S.V.L.T.; visualization, L.M.C.A., A.S.S.C. and S.V.L.T.; supervision, M.O.R.S., I.C.G. and A.V.M.G.; project administration, M.O.R.S. and I.C.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study involved the administration of an anonymized survey instrument to civil servants who voluntarily participated, with no intervention and no access to sensitive personal data. No work was done with vulnerable populations. In accordance with the UTP Code of Ethics for Researchers and Scientific Integrity (INV-RG002, version 05), ethical approval was not required, given the non-sensitive and voluntary nature of the survey data collected.

Informed Consent Statement: The study involved the administration of an anonymized survey instrument to civil servants who voluntarily participated. No personally identifiable information was collected, and no sensitive data was accessed.

Data Availability Statement: The data supporting the results of this study were not publicly available due to confidentiality restrictions and access conditions established by the participating organizations. Access could be evaluated upon reasonable request to the corresponding author, subject to authorization from the institutions holding the information and compliance with applicable data protection regulations.

Acknowledgments: We would like to thank the participating organizations and institutions for facilitating the administration of the survey instrument and for providing institutional authorization for the study, as well as the authorities and institutional officials who authorized the release of the information. Likewise, we acknowledge the support of the authors and administrative staff of these institutions for their collaboration in organizing, validating, and submitting the data necessary for the development of the study.

Conflicts of Interest: The authors declare that they have no conflicts of interest.

References

1. Wirtz, B.W.; Langer, P.F.; Fenner, C. Artificial Intelligence in the Public Sector—A Research Agenda. *Int. J. Public Adm.* **2021**, *44*, 1103–1128. [CrossRef]
2. Wang, C.; Teo, T.S.H.; Janssen, M.F.W.H.A. Public and private value creation using artificial intelligence: An empirical study of AI voice robot users in Chinese public sector. *Int. J. Inf. Manag.* **2021**, *61*, 102401. [CrossRef]
3. Di Vaio, A.; Hassan, R.; Alavoine, C. Data intelligence and analytics: A bibliometric analysis of human–Artificial intelligence in public sector decision-making effectiveness. *Technol. Forecast. Soc. Change* **2022**, *174*, 121201. [CrossRef]
4. Campion, A.; Hernandez, M.G.; Mikhaylov, S.J.; Esteve, M. Overcoming the Challenges of Collaboratively Adopting Artificial Intelligence in the Public Sector. *Soc. Sci. Comput. Rev.* **2022**, *40*, 462–477. [CrossRef]
5. van Noordt, C.; Misuraca, G.C. Artificial intelligence for the public sector: Results of landscaping the use of AI in government across the European Union. *Gov. Inf. Q.* **2022**, *39*, 101714. [CrossRef]
6. Sharma, M.; Luthra, S.; Joshi, S.; Kumar, A. Implementing challenges of artificial intelligence: Evidence from public manufacturing sector of an emerging economy. *Gov. Inf. Q.* **2022**, *39*, 101624. [CrossRef]
7. Ishengoma, F.; Shao, D.; Alexopoulos, C.; Saxena, S.; Nikiforova, A. Integration of artificial intelligence of things (AIoT) in the public sector: Drivers, barriers and future research agenda. *Digit. Policy Regul. Gov.* **2022**, *24*, 449–462. [CrossRef]
8. Presidency of the Council of Ministers. *Catalog of Artificial Intelligence Applications in the Peruvian Government*; PCM: Lima, Peru, 2025. Available online: https://www.gob.pe/institucion/pcm/informes-publicaciones/6879780-catalogo-de-aplicaciones-con-inteligencia-artificial-en-el-estado-peruano (accessed on 6 March 2026).
9. Chen, Y.; Ahn, M.J.; Wang, Y. Artificial Intelligence and Public Values: Value Impacts and Governance in the Public Sector. *Sustainability* **2023**, *15*, 4796. [CrossRef]
10. Maragno, G.; Tangi, L.; Gastaldi, L.; Benedetti, M. Exploring the factors, affordances and constraints outlining the implementation of Artificial Intelligence in public sector organizations. *Int. J. Inf. Manag.* **2023**, *73*, 102686. [CrossRef]
11. Bian, X.; Wang, B. Exploring the Influence of Artificial Intelligence Usage on Ethical Decision Making Among Public Sector Employees: Insights into Moral Identity and Service Motivation. *Bus. Ethics Leadersh.* **2024**, *8*, 133–150. [CrossRef]
12. El Khatib, M.M.; Ahmed, G. Achieving excellence in business practices through artificial intelligence: A case study of the Dubai public sector. *Int. J. Public Sect. Perform. Manag.* **2024**, *14*, 262–277. [CrossRef]
13. Giraldi, L.; Rossi, L.; Rudawska, E.D. Evaluating public sector employee perceptions towards artificial intelligence and generative artificial intelligence integration. *J. Inf. Sci.* **2024**, 01655515241293775. [CrossRef]
14. Criado, J.I. Artificial Intelligence in the Latin American Public Sector. A comparative study based on the Ibero-American Charter on Artificial Intelligence in Public Administration; Inteligencia Artificial en el Sector Público latinoamericano. Estudio Comparado a partir de la Carta Iberoamericana de Inteligencia Artificial en la Administración Pública. *Reforma Y Democr.* **2024**, *88*, 116–143. [CrossRef]
15. Al Wael, H.; Abdallah, W.; Ghura, H.; Buallay, A.M. Factors influencing artificial intelligence adoption in the accounting profession: The case of public sector in Kuwait. *Compet. Rev.* **2024**, *34*, 3–27. [CrossRef]
16. Mellouli, S.; Janssen, M.F.W.H.A.; Ojo, A.K. Introduction to the Issue on Artificial Intelligence in the Public Sector: Risks and Benefits of AI for Governments. *Digit. Gov. Res. Pract.* **2024**, *5*, 1–6. [CrossRef]
17. Alshehhi, K.; Cheaitou, A.; Rashid, H.S.J. Procurement of Artificial Intelligence Systems in UAE Public Sectors: An Interpretive Structural Modeling of Critical Success Factors. *Sustainability* **2024**, *16*, 7724. [CrossRef]
18. Dreyling, R.; Lemmik, J.; Tammet, T.; Pappel, I. An Artificial Intelligence Maturity Model for the Public Sector: A Design Science Approach. *TalTech J. Eur. Stud.* **2024**, *14*, 217–239. [CrossRef]
19. Alrawahna, A.S.; Alzghoul, A.; Awad, H.A.H. The Impact of Artificial Intelligence on Public Sector Decision-Making: Benefits, Challenges, and Policy Implications. *Int. Rev. Manag. Mark.* **2025**, *15*, 125–138. [CrossRef]
20. Rakšnys, A.V.; Gudelis, D.; Guogis, A. The Uses of Artificial Intelligence in the Public Sector: Challenges and Prospects; Dirbtinio Intelekto Panaudojimas Viešajame Sektoriuje: Iššūkiai ir Perspektyvos. *Public Policy Adm.* **2025**, *24*, 467–477.
21. Chaniago, H.; Hidayat, H.; Efawati, Y. Intrinsic Motivation and the Use of Artificial Intelligence (AI) in the Public Sector: Evidence from Indonesia; Motivação Intrínseca e o Uso da Inteligência Artificial (IA) no Setor Público: Evidências da Indonésia. *Rev. Bras. Politicas Publicas* **2025**, *15*, 412–427.
22. Liu, H.K.H.; Tang, M.-C.; Collard, A.S.J. Hybrid intelligence for the public sector: A bibliometric analysis of artificial intelligence and crowd intelligence. *Gov. Inf. Q.* **2025**, *42*, 102006. [CrossRef]
23. Al-Saba, R.Q.M. Evaluation the impact of artificial intelligence (AI) on the implementation of merit criteria in employment in the public sector. *Transform. Gov. People Process Policy* **2025**, *19*, 414–427. [CrossRef]
24. Abdallah, W.; Harraf, A.; Al Wael, H. Factors influencing artificial intelligence implementation in the accounting industry: A comparative study among private and public sectors. *J. Financ. Report. Account.* **2025**, *23*, 1509–1530. [CrossRef]

25. Panda, M.; Hossain, M.M.; Puri, R.; Ahmad, A. Artificial intelligence in action: Shaping the future of public sector. *Digit. Policy Regul. Gov.* **2025**, *27*, 668–686. [CrossRef]
26. Islam, M.M.; Tareque, M. Public sector innovation outcome-driven sustainable development in Bangladesh: Applying the dynamic autoregressive distributed lag simulations and Kernel-based regularised least square machine learning algorithm approaches. *J. Public Policy* **2023**, *43*, 326–357. [CrossRef]
27. Sundberg, L.T.; Holmström, J. Fusing domain knowledge with machine learning: A public sector perspective. *J. Strateg. Inf. Syst.* **2024**, *33*, 101848. [CrossRef]
28. Delfos, J.; Zuiderwijk, A.M.G.; van Cranenburgh, S.; Chorus, C.G.; Dobbe, R.I.J. Integral system safety for machine learning in the public sector: An empirical account. *Gov. Inf. Q.* **2024**, *41*, 101963. [CrossRef]
29. Dunleavy, P.; Margetts, H.; Bastow, S.; Tinkler, J. New Public Management Is Dead—Long Live Digital-Era Governance. *J. Public Adm. Res. Theory* **2006**, *16*, 467–494. [CrossRef]
30. Potoski, M.; Lund-Sørensen, B.; Petersen, O.H.H. Measuring Transaction Costs in Public Sector Contracting Through Machine Learning and Contract Text. *Public Adm. Rev.* **2025**, *86*, 199–216. [CrossRef]
31. Simmonds, H.; Gazley, A.J.; Kaartemo, V.; Renton, M.; Hooper, V.A. Mechanisms of service ecosystem emergence: Exploring the case of public sector digital transformation. *J. Bus. Res.* **2021**, *137*, 100–115. [CrossRef]
32. Saxena, D.K.; Muzellec, L.; McDonagh, J.J. From Bureaucracy to Citizen-Centricity: How the Citizen-Journey Should Inform the Digital Transformation of Public Services. *Int. J. Electron. Gov. Res.* **2022**, *18*, 1–17. [CrossRef]
33. Wang, C.; Ma, L. Digital transformation of citizens' evaluations of public service delivery: Evidence from China. *Glob. Public Policy Gov.* **2022**, *2*, 477–497. [CrossRef]
34. Shibambu, A.; Ngoepe, M. Enhancing service delivery through digital transformation in the public sector in South Africa. *Glob. Knowl. Mem. Commun.* **2024**, *74*, 63–76. [CrossRef]
35. Savignon, A.B.; Zecchinelli, R.; Costumato, L.; Scalabrini, F. Automation in public sector jobs and services: A framework to analyze public digital transformation's impact in a data-constrained environment. *Transform. Gov. People Process Policy* **2024**, *18*, 49–70. [CrossRef]
36. Afzal, M.; Panagiotopoulos, P. Coping with digital transformation in frontline public services: A study of user adaptation in policing. *Gov. Inf. Q.* **2024**, *41*, 101977. [CrossRef]
37. Sigurjonsson, T.O.O.; Jónsson, E.; Guðmundsdóttir, S. Sustainability of Digital Initiatives in Public Services in Digital Transformation of Local Government: Insights and Implications. *Sustainability* **2024**, *16*, 10827. [CrossRef]
38. Szedmák, B.; Varga, L.; Szabó, R.Z. Digital Transformation of Public Services: The Case of the Document Management Application. *Int. J. Public Adm.* **2025**, *2*, 1–18. [CrossRef]
39. Hossain, M.A.; Akter, S.; Dwivedi, Y.K.; Maier, C.; Janssen, M.; Rana, N.P.; Currie, W. Digital Transformation Empowerment Capabilities in Public Service Systems. *J. Comput. Inf. Syst.* **2025**, 1–23. [CrossRef]
40. Mustofa, A.; Haryati, E.; Ismail, S.B. The Impact of Digital Transformation on Public Services Governance: A Quality Assessment Scale Approach in Urban Municipalities. *J. Gov. Regul.* **2025**, *14*, 156–165. [CrossRef]
41. Irfan, M.; Haryono, D. Digital transformation in public service delivery in Palu city: A study on the digitalization of public services through the 'SanguPalu' application. *Int. J. Public Policy Adm. Res.* **2025**, *12*, 229–242. [CrossRef]
42. Djatmiko, G.H.; Sinaga, O.; Pawirosumarto, S. Digital Transformation and Social Inclusion in Public Services: A Qualitative Analysis of E-Government Adoption for Marginalized Communities in Sustainable Governance. *Sustainability* **2025**, *17*, 2908. [CrossRef]
43. Pu, S.; Ou, Y.; Bai, O. Government Public Services and Regional Digital Transformation for Sustainable Development: An Innovation Ecosystem Perspective. *Sustainability* **2025**, *17*, 5314. [CrossRef]
44. Sacchi, S.; Scarano, G. Digital transformation of public employment services in the post-pandemic era. Evidence from Italy as a latecomer country. *Aust. J. Soc. Issues* **2025**, *60*, 456–472. [CrossRef]
45. Valaskova, K.; Nagy, M.; Figura, M.; Rousek, P. Res publica digitalis: The uneven digital transformation of the European public sector and the impact of policy disparities on governance, service efficiency, and socioeconomic inclusion. *Oeconomia Copernic.* **2025**, *16*, 1177–1260. [CrossRef]
46. Mohamed, A.M.A. Digital Transformation of Public Utilities A Comparative Study on the Adaptation of Digital Services in Egypt, France, and the United Kingdom; La Transformation Numérique Des Services Publics: Étude Comparative Sur L'adaptation Des Services Numériques En Égypte, En France Et Au Royaume-Uni; A Transformação Digital Dos Serviços Públicos: Um Estudo Comparativo Sobre A Adaptação Dos Serviços Digitais No Egito, Na França E No Reino Unido. *Rev. Jurid.* **2025**, *4*, 15–34.
47. Huang, X.; Lan, F. Navigating Digital Transformation: Synergistic Pathways and Challenges in County-level Public Service Development. *Sage Open* **2025**, *15*, 21582440251. [CrossRef]
48. David, S.; Vîrlănuţă, F.O.; Bacalum, S.; Bărbuţă-Mişu, N.; Mihai, I.O. Exploring Digital Transformation as a Catalyst for Institutional Agility in the Delivery of Public Services. *Can. J. Adm. Sci.* **2025**, *42*, 514–532. [CrossRef]

49. Alzebda, S.A.; Matar, M.A.I. Factors affecting citizen intention toward AI acceptance and adoption: The moderating role of government regulations. *Compet. Rev.* **2025**, *35*, 434–455. [CrossRef]
50. Rathnayake, A.S.; Nguyen, T.D.H.N.; Ahn, Y. Factors Influencing AI Chatbot Adoption in Government Administration: A Case Study of Sri Lanka's Digital Government. *Adm. Sci.* **2025**, *15*, 157. [CrossRef]
51. Omonov, M.S.; Ahn, Y. Towards Smart Public Administration: A TOE-Based Empirical Study of AI Chatbot Adoption in a Transitioning Government Context. *Adm. Sci.* **2025**, *15*, 324. [CrossRef]
52. Gao, Q.; Wang, S.; Liang, Z.; Wang, G.; Guo, L. Sailing with the cultural winds: The impact of National culture on government AI adoption. *Glob. Public Policy Gov.* **2025**, *5*, 251–273. [CrossRef]
53. Cho, S.; Hur, J.; Kim, D. Bridging trust in AI and its adoption: The role of organizational support in AI chatbot implementation in korean government agencies. *Gov. Inf. Q.* **2025**, *42*, 102081. [CrossRef]
54. Rodriguez-Saavedra, M.O.; Benavides, L.G.B.; Galindo, I.C.; Ascuña, L.M.C.; Gonzales, A.V.M.; Lopez, J.W.M.; Arguedas-Catasi, R.W. Augmented Reality as an Educational Tool: Transforming Teaching in the Digital Age. *Information* **2025**, *16*, 372. [CrossRef]
55. Rodriguez-Saavedra, M.O.; Prado, E.A.D.; Sarolli, A.E.D. From Getting Out to Vote: Digital Political Narratives and Their Effect on Young People's Electoral Decision-Making. *Vis. Rev. Int. Vis. Cult. Rev./Rev. Int. Cult.* **2026**, *18*, 241–260. [CrossRef]
56. Babšek, M.; Ravšelj, D.; Umek, L.; Aristovnik, A. Artificial Intelligence Adoption in Public Administration: An Overview of Top-Cited Articles and Practical Applications. *AI* **2025**, *6*, 44. [CrossRef]
57. Criado, J.I.; Sandoval-Almazán, R.; Gil-Garcia, J.R. Artificial Intelligence and Public Administration: Understanding Actors, Governance, and Policy from Micro, Meso, and Macro Perspectives. *Public Policy Adm.* **2025**, *40*, 173–184. [CrossRef]
58. Busuioc, M. Accountable Artificial Intelligence: Holding Algorithms to Account. *Public Adm. Rev.* **2021**, *81*, 825–836. [CrossRef]
59. Nam, J.; Bell, E. Efficiency or Equity? How Public Values Shape Bureaucrats' Willingness to Use Artificial Intelligence to Reduce Administrative Burdens. *Public Perform. Manag. Rev.* **2024**, *48*, 1–34. [CrossRef]
60. Chen, T.; Gasco-Hernandez, M. Uncovering the Results of AI Chatbot Use in the Public Sector: Evidence from US State Governments. *Public Perform. Manag. Rev.* **2024**, *48*, 1–26. [CrossRef]
61. Shin, B. Exploring the Potential of Machine Learning to Reduce Administrative Burden in Participatory Budgeting: A Case Study of Seoul. *J. Public Budg. Account. Financ. Manag.* **2026**, *38*, 237–264. [CrossRef]
62. Tofan, D.C. Integrating Artificial Intelligence into Public Administration: Challenges and Vulnerabilities. *Adm. Sci.* **2025**, *15*, 149. [CrossRef]
63. Li, Z. Quantitative Analysis of Satisfaction with Chinese Local Government Digital Public Service Policies Using XGBoost Algorithm. *Systems* **2025**, *13*, 808. [CrossRef]
64. Silalahi, A.D.K.; Tsai, J.-C.; Wang, Y.-H. Decoding Effectiveness and Efficiency in AI-Enabled Public Services: A Configurational Pathway to Citizen and Employee Satisfaction. *Front. Polit. Sci.* **2025**, *7*, 1560180. [CrossRef]
65. Medaglia, R.; Gil-Garcia, J.R.; Pardo, T.A. Artificial Intelligence in Government: Taking Stock and Moving Forward. *Soc. Sci. Comput. Rev.* **2023**, *41*, 123–140. [CrossRef]
66. Madan, R.; Ashok, M. AI Adoption and Diffusion in Public Administration: A Systematic Literature Review and Future Research Agenda. *Gov. Inf. Q.* **2022**, *39*, 101726. [CrossRef]
67. Neumann, O.; Guirguis, K.; Steiner, R. Exploring Artificial Intelligence Adoption in Public Organizations: A Comparative Case Study. *Public Manag. Rev.* **2024**, *26*, 114–141. [CrossRef]
68. Lundberg, S.M.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 4765–4774. [CrossRef]
69. Rodriguez-Saavedra, M.O.; López, R.E.G.; Velazco, R.P.; Gonzales, H.E.A.; Pérez, A.B.D.; Apaza, O.A.; Pajuelo, J.A.E.; González, R.A.P.; Galindo, I.C.; Ascuña, L.M.C.; et al. Political Science and Governance: Citizen Participation and Rebuilding Trust in the State. *Soc. Sci.* **2026**, *15*, 1. [CrossRef]