
| RESEARCH ARTICLE

Delivering Smarter: How Artificial Intelligence Transforms Organizational Performance and Service Quality in a Public Postal Operator

Chiraz Rzaïgui¹, Imen Chaouch², and Rym Khanfir³✉

¹ *Technologue agrégée en économie et gestion, Institut Supérieur des Etudes Technologiques de Bizerte Tunisie*

rzaigui_chiraz@yahoo.fr

² *University of Tunis, Tunisia* imenzarroukch@gmail.com

³ *Laboratoire RIGUEUR, University of Tunis Tunisia* rymkhanfir4@gmail.com

Corresponding Author: Rym Khanfir **E-mail:** rymkhanfir4@gmail.com

| ABSTRACT

In the era of digital transformation, artificial intelligence (AI) is emerging as a strategic lever for modernizing public postal operators, which have historically struggled with counter congestion and rising service-quality expectations. This research investigates the extent to which AI integration can improve organizational performance and service quality within the Tunisian Post. A qualitative triangulation strategy was used, combining fifteen semi-structured interviews with managers and staff and a netnographic study of 25 Google Maps reviews collected between May 2025 and April 2026. Convergent findings reveal recurring and clearly identified irritants on the users' side, spontaneously expressed in online reviews, concentrated on waiting times, parcel-tracking reliability and shipment security. Qualitative interviews confirm an overall positive perception of AI's potential contribution to the operational and economic dimensions of performance, but a more modest perception of its strategic dimension. The netnographic study corroborates these findings, revealing recurring irritants linked to service slowness, technical outages and opening hours. Based on these results, operational recommendations are proposed, including a smart queue-management system, predictive parcel tracking and a Tunisian-dialect conversational agent. Theoretically, the study argues that organizational readiness and public legitimacy for AI adoption operate as interacting, rather independent, conditions, and that internal optimism about AI's operational benefits does not by itself secure the external trust on which its acceptability depends. The article discusses the theoretical and managerial implications of these findings, positions the contribution within the literature on AI adoption in public and postal services, and identifies the study's methodological limitations along with avenues for future research.

| KEYWORDS

Artificial intelligence; organizational performance; service quality; postal sector; public service; qualitative triangulation; netnography; Tunisia

| ARTICLE INFORMATION

ACCEPTED: 01 June 2026

PUBLISHED: 08 July 2026

DOI: 10.32996/jpar.2026.3.1.1

1.Introduction

More than a thousand offices, nearly nine thousand employees, and a mission of financial inclusion that no other player fulfills on the same scale: the Tunisian Post has occupied, since its creation in 1999, a place that its status as a public establishment of an industrial and commercial nature is not enough to summarize. However, this same institution faces very concrete and daily difficulties, overcrowding at counters, variable processing times, improved reliability of shipment tracking, which call into question its ability to honor its universal service mission in a context of accelerated digitalization and growing competition from private delivery operators.

Faced with these difficulties, artificial intelligence (AI) is gradually establishing itself, in literature as well as institutional discourse, as a lever for modernization for organizations, including public institutions (Madan & Ashok, 2023). In the postal sector specifically, recent work documents the growing use of AI for automated sorting, address recognition, route optimization and demand forecasting (Raizonville, 2024; Universal Postal Union, 2023). However, this literature emphasizes that performance gains are not automatic: they depend on internal organizational capacities, the quality of the available data and especially on the acceptability of the tools by agents and by users themselves (Madan & Ashok, 2023; Schiff, Schiff & Pierson, 2022).

However, this acceptability remains poorly documented empirically in the postal context, and more broadly in the Maghreb countries: the existing literature mainly focuses on Western contexts or remains focused on a prescriptive discourse on the expected benefits of AI, without always confronting this theoretical potential with the concrete experience of professionals and users of a given postal operator. Few works intersect the internal gaze of a postal institution with the spontaneous, unsolicited expression of its own users, and this contextual and methodological gap is symptomatic of a more fundamental theoretical one: the literature on AI adoption in public organizations tends to evaluate organizational readiness (Madan & Ashok, 2023; Neumann, Guirguis & Steiner, 2024) and citizen trust or public value (Twizeyimana & Andersson, 2019; Schiff, Schiff & Pierson, 2022) as parallel, largely separate conditions for successful adoption, rather than examining how they interact within a single organization, and whether favorable internal readiness can coexist with, or even mask, a deficit of external legitimacy. It is this unresolved question, of which the contextual and methodological gap identified above is only the empirical symptom, that underpins the problem of this article: to what extent can the integration of artificial intelligence constitute a lever for improving organizational performance and quality of service within a public postal operator such as the Tunisian Post, and what does the relationship between internal and external perceptions of this potential reveal about the conditions of AI adoption in captive public services more generally?

This issue can be broken down into two complementary research questions: (1) How do the institution's professionals perceive the potential contribution of AI on different dimensions of organizational performance? (2) What are the priority irritants expressed, spontaneously, by the users themselves, and to what extent do they overlap with the diagnosis made by the professionals?

The objective of this research is to draw up, through a qualitative triangulation combining semi-structured interviews and netnography, an empirical diagnosis of the priorities and obstacles associated with the integration of AI within the Tunisian Post, in order to draw operational recommendations based on a dual reading, internal and external, of the postal field.

On the theoretical level, this research mobilizes and articulates three corpora generally treated separately, the adoption of AI in public organizations, multidimensional organizational performance and quality of service in public services, not merely to apply them jointly to the still poorly documented case of the postal sector in the North African context, but to examine, through the empirical relationship between internal (professional) and external (user) perceptions, whether organizational readiness and public legitimacy for AI adoption should be treated as interacting rather than independent conditions. On the managerial level, it aims to equip the institution's decision-makers with a prioritized diagnosis, directly based on the field, to guide digital investment choices rather than relying on a generic discourse on the supposed benefits of AI.

The article is structured as follows: Section 2 presents the theoretical framework and positions the contribution in relation to the existing literature; Section 3 details the methodology; Section 4 presents the results; Section 5 discusses these results in the light of the literature and draws theoretical and managerial implications; Section 6 concludes by specifying the limitations of the study and avenues for future research.

2. Literature review and conceptual framework

2.1. Artificial intelligence in public and postal organizations

Artificial intelligence covers a set of systems capable of simulating human cognitive functions such as learning, reasoning and decision-making (Russell & Norvig, 2020). From the expert systems of the 1980s to contemporary deep learning (Goodfellow, Bengio & Courville, 2016), its scope has expanded considerably, now affecting logistics, customer relations and decision support. In services in general, a recent theory distinguishes four forms of intelligence that can be mobilized by AI, mechanical, analytical, intuitive, and empathetic, and highlights that tasks with a strong relational component remain, to date, the most difficult to automate (Huang & Rust, 2018). In the postal sector specifically, recent literature documents an increasing use of AI for automated sorting, address recognition, route optimization and demand forecasting (Raizonville, 2024; Universal Postal Union, 2023). However, this work emphasizes that performance gains are not automatic: they depend heavily on internal organizational capacities, the quality of the available data, and the acceptability of the tools by agents and users, an acceptability that is itself conditioned, on the part of agents, by the fear that their workstation will be replaced by these same technologies (Brougham & Haar, 2018).

In the broader field of public administration, a European mapping of the use of AI in public services confirms a still uneven diffusion across administrations and countries (Misuraca, van Noordt & Boukli, 2020), and a recent systematic review notes that the adoption of AI remains heterogeneous and faces organizational, regulatory and cultural barriers that go beyond the mere availability of technology (Madan & Ashok, 2023), a finding corroborated by a comparative case study on the adoption of AI in public organizations (Neumann, Guirguis & Steiner, 2024) and by an OECD review (2024) on government readiness for AI. More broadly, the literature on the public value of digital technology shows that the benefits of the digitalization of administrations are not limited to operational efficiency but also affect citizens' trust in the institution (Twizeyimana & Andersson, 2019). Other studies warn of the risk of a "failure of public value", when AI systems are deployed without a clear articulation with public service purposes (Schiff, Schiff & Pierson, 2022), against broader negative effects of AI in public action (Valle-Cruz, García-Contreras & Gil-García, 2024), or point out that users tend to generalize a one-off failure of AI to all public action, a phenomenon described as "algorithmic transfer" (Longoni, Cian & Kyung, 2023). These observations suggest that AI should not be considered as a neutral technical solution, but as an organizational object whose appropriation by internal (agents) and external (users) stakeholders is a condition for success.

2.2. Organizational performance and quality of service in public services

Organizational performance is understood here, in accordance with a classic conception, as the ability of an organization to achieve its objectives while optimizing the use of its resources (Kaplan & Norton, 1996). This performance is broken down into several complementary dimensions: economic (cost control, value creation), operational (efficiency of daily processes), social (working conditions, equity) and strategic (ability to adapt and innovate in the long term). This multidimensional reading, rather than an unequivocal focus on profitability, is particularly suited to public organizations, whose mission and performance criteria differ from those of private organizations (Boyne, 2002). However, the application of the Balanced Scorecard to public organizations is not mechanical: it requires adapting the four dimensions to the constraints specific to the public sector, in particular the multiplicity of stakeholders and the difficulty of translating the public service mission into financial indicators (Northcott & Taulapapa, 2012).

Quality of service, on the other hand, is classically understood through the SERVQUAL model, which defines it as the gap between users' expectations and their perception of the service received, measured on dimensions such as reliability, responsiveness, assurance, empathy and tangible elements (Parasuraman, Zeithaml & Berry, 1988). Applied to public services, this framework must nevertheless be adapted: users of a public postal service generally do not have a choice of operator, which changes the nature of the service relationship and increases the weight of the dimensions related to accessibility, fair treatment and predictability of deadlines (Wisniewski, 2001). It is this contextual adaptation that guided the construction of the collection tools used in this study.

2.3. Conceptual framework of the study

The conceptual framework of the study is explicitly based on three theoretical anchors, each of which sheds light on one of the three interacting poles of the model. Organizational performance is structured according to the logic of the Balanced Scorecard (Kaplan & Norton, 1996), which invites us to go beyond the sole financial reading in favor of a multidimensional reading articulating the economic, operational, social and strategic dimensions, a grid directly used in the interview guide to structure the collection of perceptions of the institution's professionals. The quality of service perceived by the user is apprehended through the SERVQUAL model (Parasuraman, Zeithaml & Berry, 1988), adapted to the context of a captive public service where accessibility, reliability and predictability of deadlines weigh more heavily than in a competitive market context. Finally, the acceptance of AI itself, both on the side of agents and users, is informed by the unified theory of the acceptance and use of technology, UTAUT (Venkatesh, Morris, Davis & Davis, 2003), whose four determinants, performance expectancy, effort expectancy, social influence and facilitating conditions, offer a relevant reading grid for interpreting, in section 5, the relationship between the perceived openness to AI and the reliability of the technical infrastructures already in place.

The study does not test these theories in the form of a causal model formalized by statistical hypotheses; it mobilizes them as reading grids to structure the collection (Balanced Scorecard, SERVQUAL) and interpret the results (UTAUT), in an exploratory and diagnostic posture consistent with the still emerging nature of the subject in the North African postal context and with the methodological choice of a qualitative triangulation.

3. Methodology

3.1. Research design: a two-pronged qualitative triangulation

To answer the research question, the study uses a qualitative triangulation strategy, combining two independent and complementary sources: semi-structured interviews with professionals of the institution, and a netnographic study of the opinions published online by users. This triangulation aims to strengthen the validity of the findings by cross-referencing solicited declarative data (interviews) and unsolicited spontaneous data (online opinions), reducing the risk of social desirability

bias specific to purely declarative methods and allowing the internal gaze of professionals to dialogue with the spontaneous expression of users.

3.2. Qualitative component: semi-structured interviews

This component is based on fifteen semi-structured interviews conducted with managers and agents of the Tunisian Post holding different functions (office manager, section manager, counter and loading cabin agents), to collect a plurality of internal points of view on the potential contribution of AI.

Table 1. Respondent profile (n=15)

Function	Workforce (n)	Average seniority
Office Manager	3	14 years old
Section Leader	4	11 years old
Counter Agent	5	7 years
Loading cabin agent	3	9 years
Total	15	—

A structured interview guide focused the discussions on six themes: the current situation of the service, the current use of AI tools, and the perceived impact of AI on economic, operational, strategic and social performance. Lasting an average of 57 minutes, the interviews were conducted face-to-face at the respondents' workplace, recorded and then fully transcribed before coding. The verbatims collected were subjected to a thematic content analysis, using a coding grid identifying the frequency of appearance of each unit of meaning per respondent, in line with the principles of thematic analysis in qualitative research (Braun & Clarke, 2006), and their subsequent updating on the reflective practice of this method (Braun & Clarke, 2019). To guarantee the reliability of the coding, a double coding independent of the corpus was carried out by two coders (Coder 1 and Coder 2). Inter-coder agreement, measured by Cohen's Kappa coefficient, amounted to $\kappa = 0.84$, a level of agreement considered very satisfactory regarding the usual thresholds of interpretation (Landis & Koch, 1977); Residual differences were resolved by discussion until consensus was reached.

Thematic saturation was monitored interview by interview, by counting the number of new themes that appeared in the coding grid with each new interview conducted, in accordance with the protocol popularized by Guest, Bunce and Johnson (2006). No additional themes appeared from interview 12 onwards: the last three interviews (13 to 15) only made it possible to find themes that had already been identified, without bringing out new ones. These three interviews were nevertheless kept in the final corpus, in accordance with the practice of confirming saturation by conducting a few additional interviews after the apparent inflection point, rather than stopping collection at this point. This so-called "code saturation", which guarantees that all the themes have been identified, must nevertheless be distinguished from the more demanding "meaning" saturation, which requires an in-depth understanding of each theme and generally requires a higher number of interviews (Hennink, Kaiser & Marconi, 2017); This point is taken up as a methodological limitation in section 6.

Table 2. Progression of thematic saturation

Interviews	New themes identified	Observation
1 to 5	6	Rapid emergence of key themes (operational, economic, social performance)
6 to 9	3	Emergence of secondary themes (strategic dimension, contextual factors)
10 to 12	1	Last marginal theme identified in interview n°12
13 to 15	0	No new themes: saturation confirmed

The theoretical anchoring of the interview guide stems directly from the conceptual framework presented in section 2.3: themes 3 to 6 take up the four dimensions of the Balanced Scorecard as they are (Kaplan & Norton, 1996), theme 1 mobilizes the dimensions of the SERVQUAL model relating to the reliability and responsiveness of the service (Parasuraman, Zeithaml & Berry, 1988), and Theme 2 is based on the determinants of technological acceptance of the UTAUT model, perceived utility, perceived ease of use and enabling conditions (Venkatesh, Morris, Davis & Davis, 2003). The sub-themes were derived from the literature on AI in public and postal organizations mobilized in section 2.1.

Table 3. Interview guide: themes, sub-themes and theoretical anchoring

No.	Theme	Sub-themes (from the literature)	Theoretical anchoring
1	Current status of the service	Overcrowding of the counters; processing times; Reliable tracking	SERVQUAL (Parasuraman, Zeithaml & Berry, 1988)
2	Current use of AI tools	Tools already in place; degree of familiarity; Perceived obstacles to use	UTAUT (Venkatesh, Morris, Davis & Davis, 2003)
3	Perceived impact, economic performance	Cost control; Value creation	Balanced Scorecard (Kaplan & Norton, 1996)
4	Perceived impact, operational performance	Process efficiency; Reduced errors. Processing speed	Balanced Scorecard (Kaplan & Norton, 1996)
5	Perceived impact, strategic performance	Decision support; innovation; long-term adaptation	Balanced Scorecard (Kaplan & Norton, 1996)
6	Perceived impact, social performance	Working conditions; equity; Fear of job substitution	Balanced Scorecard (Kaplan & Norton, 1996); STARA (Brougham & Haar, 2018)

The size of this sample ($n = 15$), while it makes it possible to achieve a form of thematic saturation on the main findings, must nevertheless be stressed as an assumed methodological limitation: it allows neither a statistical generalization, nor a claim to the representativeness of all the institution's staff. On the other hand, it offers a contextualized qualitative perspective, consistent with the exploratory vocation of this component within the triangulation system, and whose value lies precisely in its comparison with the other source of data.

3.3. Netnographic component: e-reputation on Google Maps

The second component uses netnography, a qualitative method consisting of analyzing the digital traces left spontaneously by users in an environment not controlled by the researcher (Kozinets, 2002; Kozinets & Gretzel, 2024). Unlike the interview, this approach captures a spontaneous, unsolicited expression that is less exposed to social desirability bias. The choice was made of a non-participatory (passive) netnography, which is limited to the observation and analysis of opinions already published without the intervention of the researcher in the exchanges, a position recognized as legitimate as soon as it is explained and distinguished from a participatory netnography (Costello, McDermott & Wallace, 2017). A corpus of 25 textual notices was compiled from comments published on Google Maps concerning Tunisian Post agencies and Rapid-Poste centers, over a collection window from May 2025 to April 2026. Only reviews containing a verbatim were retained, to the exclusion of simple star ratings, to give priority to the semantic richness of the corpus. This volume, although modest, has made it possible to achieve a form of thematic saturation, with the same irritants recurring from one opinion to another. The data was coded according to an eight-axis analysis grid (staff reception and performance, waiting times and times, schedules and accessibility, logistical performance, monitoring and transparency, technical reliability, organization and procedures, overall satisfaction). The same protocol of independent double coding was applied to this corpus by the same two coders (Coder 1 and Coder 2), with a Cohen's Kappa coefficient of $\kappa = 0.87$, also very satisfactory (Landis & Koch, 1977), the occasional deviations having been discussed until consensus.

Table 4. Netnographic analysis grid: thematic axes

Axe	Contents
1. Reception and performance of staff	Friendliness, availability and responsiveness of the agents in contact with the public
2. Wait times and wait times	Waiting time at the counter and processing time of transactions
3. Schedules and accessibility	Adequacy of opening hours and physical accessibility of branches
4. Logistics performance	Fast and reliable delivery of parcels and mail
5. Monitoring and transparency	Reliability and availability of follow-up information communicated to the user
6. Technical reliability	Breakdowns, unavailability and malfunctions of information systems
7. Organization and procedures	Clarity and simplicity of administrative procedures
8. Overall satisfaction	General assessment expressed by the user in his or her opinion

Table 5. Reliability of double coding (inter-coder agreement)

Corpus	Coders	Cohen's Kappa	Interpretation (Landis & Koch, 1977)
Interviews (n = 15)	Coder 1, Coder 2	$\kappa = 0,84$	Very satisfactory agreement
Netnographic opinions (n = 25)	Coder 1, Coder 2	$\kappa = 0,87$	Very satisfactory agreement

From an ethical point of view, only public notices, already freely accessible on an open platform, have been mobilized, without the collection of personally identifiable data, in accordance with the usual recommendations for non-intrusive netnographic research (Kozinets & Gretzel, 2024).

4. Results

4.1. Results of the qualitative study

The thematic analysis of the fifteen interviews reveals an overall favorable perception of the contribution of AI but unevenly distributed according to the dimensions of performance. The frequency grid shows a high density of citations on the operational and economic dimensions of performance (respectively 15 and 14 units of meaning over the fifteen interviews), with respondents insisting on the reduction of errors, the speed of processing and the reduction in operational costs made possible by automation. The social dimension is also well represented (12 units of meaning), with agents mentioning stress reduction, greater equity in the allocation of tasks and better prevention of psychosocial risks. The strategic dimension, on the other hand, appears to be less spontaneously mobilized (10 units of meaning), with respondents referring to decision support and digital innovation in a more general and less operationalized way than the other dimensions.

Table 6. Frequency of units of meaning by performance dimension (interviews, n = 15)

Performance dimension	Units of meaning (out of 15 interviews)	Thematic illustration
Operational	15	Error reduction, speed of processing
Economical	14	Lower operational costs, optimize resources
Social	12	Stress reduction, fairness in task assignment
Strategic	10	Decision support, digital innovation

The contextual factors mentioned by all respondents (15 out of 15) concern the growing competition from private delivery operators and the rise of e-commerce, perceived as external pressures justifying an acceleration of the institution's digital transformation. This observation converges with the irritants identified by users in the netnographic component, particularly on the speed of delivery and the tracking of parcels.

A. 4.2. Results of the Netnographic Study

The analysis of the 25 Google Maps reviews corroborates and enriches the previous findings. Two categories of difficulties dominate the corpus: problems related to the behavior and availability of staff (processing times perceived as excessive, lack of responsiveness at peak times) and technical and logistical problems (recurrent breakdowns of information systems, early closures, unavailability on Saturdays, delivery incidents). These irritants directly overlap with the findings of the qualitative component, in particular waiting times, technical reliability and logistical performance, which reinforces the robustness of the diagnosis by cross-checking independent sources.

5. Discussion

5.1. Convergent triangulation

Comparing the two aspects reveals a notable convergence of diagnoses, despite distinct methods and populations. The spontaneous digital traces left by users (netnographic component) highlight a set of critical priorities: waiting time, reliability of logistics monitoring and technical availability of digital services. The institution's professionals (qualitative component), for their part, do not dispute this diagnosis; on the contrary, they identify AI as a means of responding to it, on the operational and economic dimensions of performance. This convergence between the spontaneous voice of the client and the voice of the professional constitutes an argument of reinforced validity for the proposed priorities for action, in accordance with the logic of qualitative triangulation (Kozinets & Gretzel, 2024).

Table 7. Convergence between the two components of the study

Priority identified	Qualitative component (interviews)	Netnographic Component (Online Notice)
Wait Time	Cited as an irritant to be treated as a priority by professionals	Most frequently mentioned irritating in reviews
Reliable logistics tracking	Confirmed as an area for operational improvement	Recurring complaints about package tracking and transparency
Technical reliability of digital tools	Recognized as a condition for the success of any digitalization	Recurring technical failures reported by users
Shipment security	Mentioned as an issue of trust	Delivery incidents mentioned in multiple reviews

However, one point of tension deserves to be highlighted: while professionals generally perceive AI as a credible lever, netnographic opinions reveal a punctual mistrust linked to the recurrent technical failures of the systems already in place. This finding suggests that user support for increased automation is not a priori a given and will largely depend on the perceived reliability of the tools actually deployed—a phenomenon documented as algorithm aversion, with individuals losing trust in an automated system more quickly than in a human after the same error is observed (Dietvorst, Simmons & Massey, 2015). This result is consistent with work on "algorithmic transfer", according to which users of public services tend to generalize isolated negative experiences to an entire technological device (Longoni, Cian & Kyung, 2023): the challenge is therefore not only the introduction of AI, but the quality of its ergonomics and the support of its appropriation, otherwise a first failure of use could durably damage trust in the whole of the digitalization process, a trust that, according to the literature on the subject, is gradually built and remains more fragile towards AI than towards a human interlocutor (Glikson & Woolley, 2020). Viewed through the UTAUT grid (Venkatesh, Morris, Davis & Davis, 2003), this observation suggests that facilitating conditions and effort expectancy condition, at least as much as perceived utility, the effective acceptance of AI by users.

5.2. Theoretical implications

On a theoretical level, this research advances a specific argument rather than a general illustration of AI's usefulness: it proposes that AI adoption in captive public services is conditioned less by the intrinsic performance of the technology than by a readiness asymmetry between internal and external stakeholders, an asymmetry that the convergence/divergence pattern observed here allows us to characterize precisely. Professionals and users converge on the diagnosis of the operational irritants to be addressed (waiting time, tracking, technical reliability), which would, on a purely technological reading, suggest a straightforward case for automation. Yet the same data reveal a divergence in the basis on which each group evaluates AI: professionals' reason from an internal, prospective standpoint, projecting AI onto anticipated operational and economic gains, whereas users reason from a retrospective standpoint, generalizing the reliability of past digital interactions, marked by recurring technical failures, into an anticipated judgment on future automation. This divergence is not merely descriptive: it offers an explanation, grounded in the data, for why organizational optimism about AI does not automatically translate into public trust, and it refines the algorithmic transference thesis (Longoni, Cian & Kyung, 2023) and algorithm aversion research (Dietvorst, Simmons & Massey, 2015) by locating the mechanism not in the algorithm's performance itself but in an asymmetry of reference points between the organization's internal narrative of progress and the user's cumulative experience of the institution's existing digital infrastructure. Read through this lens, the facilitating-conditions component of UTAUT (Venkatesh, Morris, Davis & Davis, 2003) should, in the specific case of captive public services, be theorized as a stakeholder-differentiated construct rather than a uniform organizational attribute, since the same infrastructure generates confidence internally while generating caution externally. This reading also speaks directly to the public-value literature: it substantiates, in a single empirical case, the more general concern that AI initiatives framed as internal performance levers can proceed without a corresponding gain, and possibly at the expense, of the external legitimacy on which public value depends (Schiff, Schiff & Pierson, 2022; Twizeyimana & Andersson, 2019). Rather than treating organizational readiness and citizen trust as two independent conditions for successful AI adoption, as much of the literature implicitly does (Madan & Ashok, 2023; Neumann, Guirguis & Steiner, 2024), the findings suggest they interact: an organization can be internally ready, in the sense of favorable professional perceptions, while remaining externally illegitimate in the eyes of users whose trust has already been eroded by prior technical failures, so that the strategic dimension of performance, the least mobilized dimension in the interviews, is precisely the one most exposed to this legitimacy gap. This internal-external readiness gap is proposed here as a concept to be tested more systematically in future comparative or quantitative research, rather than as an established finding; the qualitative and single-case nature of this study means it should be read as a plausible explanatory mechanism grounded in convergent evidence, not as a confirmed causal pathway.

5.3. Managerial implications and recommendations

From a managerial point of view, the converging results suggest that action should be prioritized around four priority axes, directly linked to the recurring irritants identified by the netnographic study and confirmed by qualitative interviews.

1) **Maximizing operational efficiency**

The implementation of an intelligent queue management system, based on peak hour forecasting algorithms to offer an appointment or a virtual ticket, would plausibly address, pending pilot evaluation, the first irritant identified in the netnographic notices, namely the waiting time at the counters. Forward-looking human resources planning, adjusted to the history of attendance, would complete this system.

2) **Logistical reliability and shipment security**

Predictive parcel tracking, capable of informing the user of a tight delivery window rather than a simple shipping status, as well as a reinforcement of automated visual control in sorting centres, would plausibly address, if implemented and subsequently evaluated, the recurring irritants identified in online notices concerning the tracking and security of parcels.

3) **Accessible digitalization and continuity of service**

Given the recurrent mentions of technical failures and difficulties of use noted in the netnographic notices, any digital solution should be designed according to the principles of simplified ergonomics and tested with non-expert users before generalization. The reinforcement of the technical infrastructure (cloud, system redundancy) is also necessary to limit the recurrent failures noted in the netnographic study.

4) **Inclusion and customer relations**

A conversational agent in Tunisian dialect, capable of processing requests for follow-up, appointment scheduling and first-level information, while directing complex requests to a human agent, would constitute a concrete and inclusive response, consistent with the overall positive openness of professionals towards AI, while preserving, for users who wish to do so, accompanied human contact.

Table 8. Summary of managerial recommendations

Priority axis	Irritant / identified need	Recommendation	Target performance dimension
1. Optimize operational efficiency	Waiting at the ticket offices	Intelligent queue management system, appointment scheduling	Operational
2. Logistical reliability and shipment security	Package Tracking & Security	Predictive parcel tracking, automated visual control	Operational, cost-effective
3. Accessible digitalization and continuity of service	Technical breakdowns, difficulty of use	Simplified ergonomics, reinforcement of the technical infrastructure	Operational, Social
4. Inclusion and customer relationship	Need for human support	Conversational agent in Tunisian dialect, orientation to a human agent	Social, strategic

These recommendations should be conceived as hypotheses for action to be tested gradually, for example in the form of evaluated pilot projects, rather than as solutions whose effectiveness would be a priori acquired: this study documents needs and priorities, but does not constitute, at this stage, an impact assessment of an AI device that has actually been deployed.

6. Limitations and Future Research Paths

Several methodological limitations must be considered in the interpretation of the results. First, the study is based exclusively on qualitative and netnographic data; it does not involve a quantitative survey to statistically measure the prevalence of irritants identified among all users of the Tunisian Post, nor to test significant differences between subgroups (age, geographical area, frequency of use). The priorities identified must therefore be read as a convergent qualitative diagnosis, and not as a statistically validated hierarchy. Secondly, the qualitative sample (n = 15) and the netnographic corpus (n = 25 opinions), although having made it possible to achieve a form of thematic saturation, remain modest in size with regard to the diversity of the institution's user and employee profiles; Their robustness is based mainly on the convergence between the two sources rather than on a claim to exhaustiveness. More precisely, the saturation observed in interview 12 corresponds to a saturation of "code" (absence of new themes), and not to the saturation of "meaning", which is more demanding, which presupposes a deep and nuanced

understanding of each theme and generally requires a higher number of interviews (Hennink, Kaiser & Marconi, 2017); The findings of the interviews should therefore be read as covering the range of topics rather than exhausting their complexity. Third, the study relies on stated perceptions (interviews) and spontaneous but non-representative expression (online reviews), and not on behavioural measures or performance indicators internal to the institution (actual deadlines, costs, error rates), which limits the scope of the findings to a perception diagnosis rather than an objective assessment of performance.

These limitations open prospects for future research. A quantitative survey conducted with a larger probability sample would make it possible to statistically measure the prevalence of the irritants identified in this study and to test comparisons between user profiles (age, geographical area, frequency of use), thus consolidating the generalization of the priorities identified. A longitudinal evaluation of an AI pilot project (e.g. the chatbot prototype mentioned in the recommendations) would offer a more direct measure of impact than the perceptual approach used here. Finally, a comparative extension to other public postal operators in the Middle East and North Africa region would make it possible to test the contextual generalization of the priorities identified in the Tunisian case.

7. Conclusion

This research examined, through a qualitative triangulation combining semi-structured interviews and a netnographic study, to what extent artificial intelligence can be a lever for improving organizational performance and quality of service at the Tunisian Post. The converging results of the two components draw up a coherent diagnosis: recurring and clearly identified irritants on the users' side (waiting times, parcel tracking, security, technical reliability), and user vigilance as to the reliability of the digital tools deployed, which will condition the acceptability of any increased automation. The institution's professionals, for their part, perceive AI as a credible lever for operational and economic improvement, rather than as a strategic lever that is immediately mobilized. On this basis, targeted operational recommendations were made, relating to intelligent queue management, predictive parcel tracking, logistics security, accessible digitalization and conversational assistance; these remain, at this stage, hypotheses for action rather than demonstrated effects, since no AI system has yet been deployed or evaluated within the institution. Beyond the Tunisian case, and beyond the operational diagnosis itself, this study suggests that organizational readiness and public legitimacy for AI adoption should be understood as interacting rather than independent conditions: an institution's internal confidence in AI's potential does not, by itself, guarantee the external trust on which the acceptability of automation ultimately depends, particularly where that trust has already been shaped by prior technical failures. This study illustrates the relevance of a qualitative triangulation approach to surface this gap, in a rigorous and objective way, before any technological deployment; testing its generality, however, will require the quantitative and longitudinal research outlined in section 6.

Funding: "This research received no external funding"

Conflicts of Interest: "The authors declare no conflict of interest."

ORCID iD (if any) Imen Chaouch ORCID: 0009-0008-7779-4167/ Rym Khanfir ORCID: 0009-0003-7627-3144

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

References

- [1] Boyne, G. A. (2002). Public and private management: What's the difference? *Journal of Management Studies*, 39(1), 97–122. <https://doi.org/10.1111/1467-6486.00284>
- [2] Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- [3] Braun, V., & Clarke, V. (2019). Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health*, 11(4), 589–597. <https://doi.org/10.1080/2159676X.2019.1628806>
- [4] Brougham, D., & Haar, J. (2018). Smart technology, artificial intelligence, robotics, and algorithms (STARA): Employees' perceptions of our future workplace. *Journal of Management & Organization*, 24(2), 239–257. <https://doi.org/10.1017/jmo.2016.55>
- [5] Costello, L., McDermott, M.-L., & Wallace, R. (2017). Netnography: Range of practices, misperceptions, and missed opportunities. *International Journal of Qualitative Methods*, 16(1), 1–12. <https://doi.org/10.1177/1609406917700647>
- [6] Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, 144(1), 114–126. <https://doi.org/10.1037/xge0000033>
- [7] Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals*, 14(2), 627–660. <https://doi.org/10.5465/annals.2018.0057>
- [8] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

- [9] Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough? An experiment with data saturation and variability. *Field Methods*, 18(1), 59–82. <https://doi.org/10.1177/1525822X05279903>
- [10] Hennink, M. M., Kaiser, B. N., & Marconi, V. C. (2017). Code saturation versus meaning saturation: How many interviews are enough? *Qualitative Health Research*, 27(4), 591–608. <https://doi.org/10.1177/1049732316665344>
- [11] Huang, M.-H., & Rust, R. T. (2018). Artificial intelligence in service. *Journal of Service Research*, 21(2), 155–172. <https://doi.org/10.1177/1094670517752459>
- [12] Kaplan, R. S., & Norton, D. P. (1996). *The balanced scorecard: Translating strategy into action*. Harvard Business School Press.
- [13] Kozinets, R. V. (2002). The field behind the screen: Using netnography for marketing research in online communities. *Journal of Marketing Research*, 39(1), 61–72. <https://doi.org/10.1509/jmkr.39.1.61.18935>
- [14] Kozinets, R. V., & Gretzel, U. (2024). Netnography evolved: New contexts, scope, procedures and sensibilities. *Annals of Tourism Research*, 104, Article 103693. <https://doi.org/10.1016/j.annals.2023.103693>
- [15] Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174. <https://doi.org/10.2307/2529310>
- [16] Longoni, C., Cian, L., & Kyung, E. J. (2023). Algorithmic transference: People overgeneralize failures of AI in the government. *Journal of Marketing Research*, 60(1), 170–188. <https://doi.org/10.1177/00222437221110139>
- [17] Madan, R., & Ashok, M. (2023). AI adoption and diffusion in public administration: A systematic literature review and future research agenda. *Government Information Quarterly*, 40(1), Article 101774. <https://doi.org/10.1016/j.giq.2022.101774>
- [18] Misuraca, G., van Noordt, C., & Boukli, A. (2020). The use of AI in public services: Results from a preliminary mapping across the EU. In *Proceedings of the 13th International Conference on Theory and Practice of Electronic Governance* (pp. 90–99). Association for Computing Machinery. <https://doi.org/10.1145/3428502.3428513>
- [19] Neumann, O., Guirguis, K., & Steiner, R. (2024). Exploring artificial intelligence adoption in public organizations: A comparative case study. *Public Management Review*, 26(1), 114–141. <https://doi.org/10.1080/14719037.2022.2048685>
- [20] Northcott, D., & Taulapapa, T. M. (2012). Using the balanced scorecard to manage performance in public sector organizations: Issues and challenges. *International Journal of Public Sector Management*, 25(3), 166–191. <https://doi.org/10.1108/09513551211224234>
- [21] OECD. (2024). *Governing artificial intelligence: Are governments ready?* OECD Publishing.
- [22] Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1988). SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality. *Journal of Retailing*, 64(1), 12–40.
- [23] Raizonville, A. (2024). The effects of artificial intelligence technologies on postal operators. In P. L. Parcu, T. Brennan, & V. Glass (Eds.), *Service challenges, business opportunities, and regulatory responses in the postal sector* (pp. 187–201). Springer. https://doi.org/10.1007/978-3-031-65599-9_13
- [24] Russell, S., & Norvig, P. (2020). *Artificial intelligence: A modern approach* (4th ed.). Pearson.
- [25] Schiff, D. S., Schiff, K. J., & Pierson, P. (2022). Assessing public value failure in government adoption of artificial intelligence. *Public Administration*, 100(3), 653–673. <https://doi.org/10.1111/padm.12742>
- [26] Twizeyimana, J. D., & Andersson, A. (2019). The public value of E-Government – A literature review. *Government Information Quarterly*, 36(2), 167–178. <https://doi.org/10.1016/j.giq.2019.01.001>
- [27] Universal Postal Union. (2023). *State of the postal sector 2023: Artificial intelligence and the dawn of hyper-collaboration*. UPU.
- [28] Valle-Cruz, D., García-Contreras, R., & Gil-García, J. R. (2024). Exploring the negative impacts of artificial intelligence in government: The dark side of intelligent algorithms and cognitive machines. *International Review of Administrative Sciences*, 90(2), 353–368. <https://doi.org/10.1177/00208523231187051>
- [29] Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27(3), 425–478. <https://doi.org/10.2307/30036540>
- [30] Wisniewski, M. (2001). Using SERVQUAL to assess customer satisfaction with public sector services. *Managing Service Quality: An International Journal*, 11(6), 380–388. <https://doi.org/10.1108/EUM0000000006279>