

Article

Institutional Readiness and the Boundaries of AI Adoption: Evidence from UAE Newsrooms

Elsayed Darwish ^{1,*}  and Ahmed Al Mansoori ² ¹ Department of Communication, College of Communication and Media Sciences, Zayed University, Dubai 19282, United Arab Emirates² Department of Media and Creative Industries, College of Humanities and Social Sciences, United Arab Emirates University, Al Ain 15551, United Arab Emirates; a.mansoori@uaeu.ac.ae

* Correspondence: elsayed.darwish@zu.ac.ae

Abstract

As Artificial Intelligence (AI) becomes embedded in newsroom operations, organizations invest in AI readiness on the assumption that preparedness travels with successful outcomes. Readiness frameworks typically aggregate diverse outcomes into composite performance measures, leaving it unclear whether readiness explains each equally well. This study examines the explanatory scope of institutional AI readiness in UAE newsrooms using a cross-sectional survey of a purposive sample of 336 media professionals. Three collection streams were harmonized and outcomes analyzed separately rather than aggregated. The seven readiness indicators showed high internal consistency ($\alpha = 0.919$) and a dominant first component accounting for 67.7% of item variance; the four conceptual domains behind the instrument were not recovered as separable dimensions, a pattern that may partly reflect a general evaluative disposition toward the organization. A one-standard-deviation increase in readiness was associated with 2.59 times higher odds of deeper adoption (95% CI [1.88, 3.56]) and with greater training provision and perceived return on AI investment, but not with practitioners' perceived quality of Arabic-language AI output, measured with a single item ($p = 0.097$). Because one readiness indicator concerns current AI use, all estimates were re-estimated without it: the odds ratio was 2.38 (95% CI [1.74, 3.25]), the organizational associations remained significant, and the association with perceived Arabic-language output quality remained non-significant ($p = 0.095$). Institutional readiness, therefore, appears to have a bounded explanatory domain in these data, showing stronger associations with organizational implementation outcomes than with perceived output quality. Because the design is cross-sectional and not all respondents answered outcome items, these are statements of association, not of effect. Studies relying on composite measures may overstate their explanatory reach.



Academic Editor: Andreu Casero-Ripollés

Received: 10 August 2026

Revised: 6 October 2026

Accepted: 8 October 2026

Published: 14 October 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.**Keywords:** artificial intelligence; newsroom adoption; institutional readiness; measurement; perceived output quality; Arabic-language technology; United Arab Emirates

1. Introduction

News organizations are investing heavily in Artificial Intelligence (AI) capability and in the organizational readiness thought to underwrite it (Beckett & Yaseen, 2023; Simon, 2024). Readiness frameworks share a premise: that governance, staff capability, resources, and implementation capacity determine how successfully an organization adopts AI (Alsheibani et al., 2018; Jöhnk et al., 2021; Holmström, 2022). The premise is rooted

in older accounts in which adoption outcomes follow from organizational characteristics rather than from the technology (Tornatzky & Fleischer, 1990; Rogers, 2003).

What remains uncertain is how far it reaches. Readiness is applied to a heterogeneous set of outcomes (deployment depth, organizational capability, governance, output quality) as though they were a single family of consequences flowing from a single organizational property. Because readiness frameworks are typically specified and scored as composites, the possibility that readiness has a bounded explanatory domain has rarely been tested directly.

The distinction matters practically: investment guided by an undifferentiated model will be well directed for some purposes and misdirected for others. Specifying the reach of a construct is a condition of using it meaningfully (Bacharach, 1989; Whetten, 1989; Busse et al., 2017).

The United Arab Emirates offers an informative setting. National AI investment has coincided with rising adoption in the media sector, while newsroom production remains heavily Arabic, a language posing distinctive computational challenges (Habash, 2010; Joshi et al., 2020). Organizational and language-dependent outcomes can therefore be observed within a single institutional context.

This study examines the explanatory scope of institutional AI readiness across a differentiated set of newsroom outcomes: deployment depth, training provision, perceived return on AI investment, and practitioners' perceived quality of Arabic-language AI output, each estimated separately rather than folded into a composite. Drawing on a cross-sectional survey of 336 media professionals, it asks whether readiness is a coherent organizational property and whether its reach is evenly distributed. Generative systems are reworking journalism's claims over authorship, verification, and professional identity (Lewis et al., 2025); knowing which of those pressures newsrooms can organize their way out of is a condition of responding to them.

The study makes two contributions. Methodologically, it tests the measurement structure of a newsroom readiness instrument prior to its use. Substantively, it asks whether a single readiness measure relates in the same way to outcomes that newsrooms arrange for themselves and to practitioners' judgments of AI output, which, in part, depend on systems outside organizational control.

2. Literature Review

2.1. Institutional Readiness and Organizational Adoption

Research on organizational technology adoption has long held that outcomes depend on organizational conditions rather than the technology alone, whether framed through technological, organizational, and environmental pressures (Tornatzky & Fleischer, 1990), as a staged process shaped by internal structure (Rogers, 2003), or through structural and managerial determinants (Damanpour, 1991).

Readiness constructs formalize the position by positing a measurable state of preparedness that precedes and predicts successful implementation, applied to inter-organizational systems (Iacovou et al., 1995), e-commerce (Molla & Licker, 2005), digital innovation (Lokuge et al., 2019), and organizational AI adoption, where readiness is decomposed into strategy, resources, culture, and data capabilities (Alsheibani et al., 2018; Jöhnk et al., 2021; Holmström, 2022). These frameworks have been validated largely against implementation outcomes. Whether readiness also accounts for downstream performance, particularly the quality of AI-generated journalistic content, has received far less attention.

Five related terms need to be kept apart. *Readiness* denotes a prior organizational condition: the resources, rules, and capacities in place before or alongside use. *Capability* denotes present operational capacity. *Adoption* denotes a behavior or stage: whether, and

how deeply, AI is used. *Implementation* refers to the process of embedding AI into routines. *Performance* denotes an outcome, here narrowed to practitioners' perceived quality of AI output. In practice, survey instruments often blur these categories, because respondents rate readiness from inside organizations where use is already underway. The instrument analyzed here does so as well, and Section 3.2.1 classifies its indicators accordingly.

2.2. Organizational Capability and Technical Capability

A parallel literature distinguishes between organizational and technical capabilities. Sociotechnical accounts hold that performance depends on the joint state of technical and social subsystems (Trist & Bamforth, 1951); later work treats technology and structure as mutually constituting (Orlikowski, 1992) and human and material agencies as imbricated (Leonardi, 2011). If the two jointly determine all outcomes, there is little reason to ask whether particular outcomes are more responsive to one than the other. In newsrooms, the question has force: an organization can restructure workflows and retrain staff, but the capabilities of AI systems remain largely exogenous to it.

Newsroom research has documented the organizational side closely: cultural inertia in legacy organizations, the absorption of innovation into existing routines, and the redistribution of editorial judgment as automated systems enter production (Thurman et al., 2017; Broussard et al., 2019; Lewis et al., 2019; Milosavljević & Vobič, 2019). Generative tools have extended this work, which now maps how practitioners appraise the hazards and possibilities of such systems (Cools & Diakopoulos, 2026) and how professional roles shape whether they are taken up (Møller et al., 2025). The technical side has served as context more often than as a variable.

2.3. AI Implementation in Multilingual and Arabic-Language News Environments

Most empirical evidence on AI in journalism comes from English-language settings (Diakopoulos, 2019; Beckett & Yaseen, 2023; Simon, 2024), including the most recent large-scale newsroom survey, which samples UK journalists (Thurman et al., 2025). In non-English environments, the readiness–quality relationship may differ, since computational resources and model capabilities are unevenly distributed (Joshi et al., 2020).

Arabic poses well-documented computational challenges: morphological complexity, orthographic ambiguity, and diglossia, in which Modern Standard Arabic and regional dialects diverge considerably (Habash, 2010). Dedicated models and benchmarks have proliferated, yet the evaluation literature reports that Arabic performance still lags behind English, attributing the gap to limited training data, dialectal diversity, and morphology (Almatham et al., 2025), with dialectal variation particularly underrepresented (Mousi et al., 2025). Benchmark progress does not establish that AI-generated Arabic reads as professionally adequate for the journalists who work with it, a distinct and open question that turns on practitioner judgment in production rather than on task performance under evaluation.

Research on Gulf media systems has expanded in parallel, examining institutional strategies, determinants of adoption, and professional attitudes (Ahmad et al., 2023; Al-Jenaibi & Bulhooon, 2024; Sonni et al., 2024; Hassouni & Mellor, 2025). The two literatures have developed separately: computational work asks what AI systems achieve in evaluation, and organizational work asks how news organizations govern adoption. Neither has examined whether journalists' evaluations of AI-generated Arabic vary with their organizations' readiness.

2.4. Measuring AI Readiness in Journalism

AI readiness instruments are typically framed as multidimensional, comprising domains such as technology, people, governance, and process, and are combined into com-

posite scores; however, their dimensional structure is usually specified on conceptual rather than empirical grounds. Where the proposed dimensions are not empirically distinguishable, reporting separate subscale scores invites conclusions that the measurement cannot support.

Established procedures for testing these assumptions include internal consistency analysis, item-level diagnostics, and dimensionality assessment by parallel analysis rather than the Kaiser criterion alone (Horn, 1965; Hayton et al., 2004). Such validation is rarely reported in journalism and media management, leaving the structural validity of many measures insufficiently examined.

2.5. *The Unresolved Boundary of Organizational Explanations*

A theoretical contribution is strengthened by specifying the conditions under which a relationship holds, and weakened when a construct travels without stated limits (Bacharach, 1989; Whetten, 1989; Busse et al., 2017). In newsroom AI research, readiness has been invoked to support adoption, capability development, governance, and output quality alike.

The gap is one of measurement and inference rather than topic. Readiness frameworks in the organizational AI literature combine strategic, resource, cultural, and data components into a single account of preparedness (Jöhnk et al., 2021; Holmström, 2022), and newsroom surveys typically report adoption, training, and perceived benefits side by side without testing whether one readiness measure relates to each in the same way (Beckett & Yaseen, 2023; Thurman et al., 2025). Little is therefore known about whether readiness explains different outcome domains equally well; when outcomes are combined, outcome-specific variation is masked rather than detected. Arabic-language AI content is an informative test case, with its quality depending on both organizational decisions (tool selection, workflow design, training) and model capabilities outside organizational control. Evaluating that reach requires validating the construct, examining distinct outcome domains within one organizational context, and estimating those relationships separately.

2.6. *Research Questions*

The research questions are organized around measurement, the two outcome classes, and exploratory heterogeneity. The descriptive state of AI integration (adoption stage, training, perceived return) is reported first as context for these questions rather than as a question in its own right.

RQ1. *What is the internal consistency and dimensional structure of the institutional AI readiness measure in newsroom data?*

RQ2. *How is institutional readiness associated with organizational implementation outcomes (adoption maturity, training provision, perceived return on AI investment, and perceived effect on professional performance), each estimated separately?*

RQ3. *How is institutional readiness associated with practitioners' perceived quality of Arabic-language AI output?*

RQ4 (exploratory). *Do institutional readiness and its associations vary by organization type, hierarchical position, or professional experience?*

RQ2 and RQ3 together address the scope question raised in Section 2.5; posed in parallel and estimated by the same procedure, they let the comparison rest on the estimates rather than the framing. RQ1 is prior to both, since a construct whose measurement prop-

erties are unestablished cannot support conclusions about reach. RQ4 is exploratory and is interpreted accordingly.

3. Method

3.1. Sampling and Data Collection Context

The study examined institutional AI readiness, AI adoption, and perceived quality of AI-generated journalistic content among media professionals in UAE news organizations. Purposive sampling recruited journalists, editors, newsroom managers, digital content specialists, and technical personnel directly involved in editorial decision-making or AI-supported workflows, across public-sector, commercial, and digital-native organizations.

Data were collected through a questionnaire administered between October 2025 and May 2026, reaching analysis as three streams: a main Arabic administration ($n = 267$), a second Arabic administration ($n = 45$), and an English administration ($n = 24$), fielded under two instrument versions differing in coding scheme and in the granularity of the experience and institution-type options. After screening for completeness and response quality, the final analytical sample comprised $N = 336$. The main Arabic stream combined 185 responses to the current instrument and 82 responses to its earlier version, collected between October 2025 and February 2026. The earlier version shared the demographic items and three of the seven readiness indicators. However, it did not include the outcome modules in their current form, so its respondents were included in the descriptive analyses but not in the readiness–outcome analyses (Section 3.6.2).

Most participants completed the questionnaire online through SurveyMonkey, using links circulated to media professionals; others completed it in face-to-face sessions, and the first author entered their answers. Because the questionnaire was distributed through professional contacts rather than to a defined sampling frame, a response rate cannot be computed. Questionnaires without substantive answers were removed during screening; partially completed questionnaires were retained, and the valid n is reported for each analysis. Records sharing an IP address (59) or an identical pattern of readiness ratings (11) were compared across all other answers; they differed on role, organization, or other items, consistent with shared workplace networks or straight-line rating, and none was treated as a duplicate. Of the 268 respondents who named their organization, the names correspond to approximately 130 organizations after merging spelling variants; the largest single organization contributed 30 respondents. Respondents from the same newsroom share policies, tools, and training, so their answers are not fully independent; Section 3.6.2 describes how this was handled.

3.2. Construct Measurement and Operationalization

Items were adapted from established technology-adoption and AI-readiness scales and reworded for newsroom workflows in bilingual publishing environments. Response formats varied by construct: the seven readiness indicators on a 0–100 metric; adoption maturity on a five-stage ordinal item; training provision, perceived return on AI investment, and perceived Arabic-language output quality as ordered categorical items; and the agreement battery on five-point Likert-type items. The questionnaire was developed in Arabic and administered in Arabic and English. The English version was translated from Arabic by the first author, who works bilingually in journalism research and teaching; no independent back-translation was conducted, which is noted as a limitation (Section 5.6). The full questionnaire in both languages is provided as Supplementary Material S8.

3.2.1. Institutional AI Readiness (RQ1)

Institutional AI Readiness was the study's primary construct: the organizational conditions that enable and govern the integration of AI into newsroom operations. Four conceptual domains guided item writing. Infrastructure and technological capacity (IR_INF) covered AI infrastructure, software access, and workflow compatibility; organizational strategy and governance (IR_GOV) covered formal strategies, governance policies, ethical guidelines, and risk-management protocols; training and skill provision (IR_TRN) covered investment in AI-related professional development; and evaluative capacity (IR_EVAL) covered the capacity to monitor implementation and assess workflow outcomes.

These domains describe the coverage the item pool was designed to achieve; they are not a factor model asserted in advance. Whether they correspond to empirically separable dimensions is treated as open and addressed in RQ1.

Respondents were asked to "move the slider to indicate the extent to which the following aspects are achieved in your media organization" (0–100). The seven indicators, translated from the Arabic instrument, were: (1) the level of actual adoption of AI technologies in daily work; (2) the degree of media and digital qualification of staff; (3) the speed of converting innovative ideas into practical applications; (4) the degree of collaboration between editorial and technical teams; (5) the extent to which audience engagement and trust are used in developing media content and services; (6) the availability of guides and guidelines on using AI and handling misleading news; and (7) reach to all audience groups (people with disabilities, older people, and speakers of different languages).

The indicators are not homogeneous in kind. Staff qualification and the availability of guidance are resources or antecedents; implementation speed, editorial–technical collaboration, and the use of audience engagement are organizational routines or processes; accessibility across audience groups is a normative commitment; and actual adoption in daily work is an implementation outcome. The first indicator was included on the premise that an organization's existing use of AI forms part of the base from which further implementation proceeds. It nevertheless overlaps conceptually with adoption maturity, one of the outcomes analyzed. All readiness–outcome estimates are, therefore, also reported for a six-indicator composite that excludes it (Section 4.9). Because every indicator is a respondent's rating of their own organization, the composite is interpreted throughout as a measure of perceived institutional readiness.

3.2.2. Internal Organizational Performance and Management Outcomes (RQ2)

Four dependent variables were operationalized: adoption depth (AD), the extent of routine AI use in editorial activities including content generation, transcription, copyediting and visual production; training provision (TP), the accessibility, quality and continuity of institutional training; perceived return on AI investment (ROI), respondents' judgment of whether their organization obtains a clear return from its investment in AI; and perceived professional performance (PP), practitioners' sense of productivity and work quality under AI integration. Each was analyzed independently rather than aggregated into a composite index.

Adoption depth was measured by the question "How do you rate your organization's level of adoption of AI technologies in media work?" with five ordered stages (Section 4.4). Training provision asked whether the respondent had received specialized AI training in their organization (adequate training; training received but insufficient; not received because the organization does not provide it; not received although available). Perceived return on AI investment was measured by the question "In your view, does your organization achieve a clear return from investment in AI technologies?", whose options included "cannot be assessed at present". Perceived professional performance asked how AI use

had affected the respondent's own work, including the option "I do not use it enough to assess the effect". In the correlational and ordinal analyses, responses that do not lie on an ordered scale ("not received although available" for training; "cannot be assessed at present" for return on investment; "I do not use it enough to assess the effect" for professional performance) and off-scale responses volunteered by respondents were treated as missing, which is why these outcomes have smaller valid n in Section 4.6.1.

3.2.3. Practitioners' Perceived Quality of Arabic-Language AI Output (RQ3)

Practitioners' perceived quality of Arabic-language AI output (hereafter, perceived Arabic-language output quality) was measured with a single item: "How do you rate the quality of AI outputs in Arabic (Modern Standard Arabic and local dialects) in your work?" (very poor, poor, average, excellent, very excellent). The item captures a general professional appraisal. It does not identify the tool or model used, the type of content, or the criterion applied (accuracy, linguistic correctness, or editorial usability), and respondents need not have observed comparable outputs. It is therefore treated as a perception shaped by tools, tasks, expectations, and professional standards, rather than as a measure of system performance. No screening question preceded it; respondents without sufficient experience could leave it blank, and 246 answered it. This distinction follows Section 2, which distinguishes between organizational readiness and the systems' technical and linguistic capabilities.

3.2.4. Control and Subgrouping Variables (RQ4)

Media organization type (print or broadcast; digital-only platform or news agency; government media office), hierarchical position (the section head or platform/project manager option, treated as leadership, versus all other roles), and professional experience were recorded to characterize the sample and provide the grouping variables for the subgroup comparisons in RQ4.

3.3. Harmonization Procedure

Mappings rested on documentary evidence rather than post hoc inference: recovered questionnaire layouts confirmed the documented options, and code-to-label co-occurrence was verified in the stream where both appeared. Where numeric codes and textual labels disagreed, textual labels were treated as authoritative and were consistently aligned with the original questionnaire wording.

- Experience was standardized into the three categories of the current instrument (<3 years, 3–10 years, >10 years); the finer bands of the earlier version (3–6, 5–10, and 7–10 years) were collapsed into the middle category.
- Organization type in the current instrument distinguishes print or broadcast outlets, digital-only platforms, news agencies, and government media offices. The earlier version's seven options were mapped onto these three (print, broadcast, and mixed print-digital to print or broadcast; digital-only platforms and news agencies to the second category; government media offices to the third), with "other" retained separately. The regression in Section 4.6.2 and the comparison in Section 4.8.1 draw only on respondents to the two Arabic administrations of the current instrument, so no cross-version mapping is applied to them.
- Categories absent from the documented questionnaires were retained separately for sensitivity analysis rather than reassigned; for one training variable, a small number of observations could not be mapped unambiguously and were excluded from analyses involving that variable only. The harmonized dataset ($N = 336$) served as the analytical dataset throughout the analysis.

3.4. Comparability-Testing Procedure

Before pooling, comparability across streams was tested. Stream sizes of 45 and 24 cannot support measurement invariance testing, so distributional comparability was assessed under a fixed-in-advance pooling rule.

For each harmonized item, response distributions were compared across versions using the chi-square test of independence, with Cramér's V for effect size and expected cell frequencies inspected for sparseness. Items showing no significant difference were pooled; where differences appeared, stream-specific results were treated as primary. Because more than 20% of expected cells fell below five for several items, exact (Fisher–Freeman–Halton) or Monte Carlo p -values were also computed (Supplementary Material S1). A non-significant test is the absence of evidence of a difference, not evidence of equivalence, and the small streams had limited power to detect real differences. Pooling is therefore treated as a working assumption, and two checks test its consequences: models adjusted for the questionnaire stream and replication in the main stream alone (Section 4.9).

3.5. Measurement Strategy

Multi-item measurement was possible only for Institutional AI Readiness. Adoption depth, training provision, perceived return on AI investment, perceived professional performance, and perceived Arabic-language output quality were each captured by a single item, as were institution type, hierarchical position, and experience; all were analyzed directly, without latent-variable specification.

The readiness composite is the mean of the available indicators for respondents who answer at least five of the seven indicators (a five-of-seven threshold, 71.4% of the item pool), a threshold that keeps each score anchored in a majority of the item pool while limiting the loss of cases due to item nonresponse. Of the 221 respondents with a composite score, 218 answered all seven indicators, and 3 answered six. A strict seven-of-seven rule is reported as a sensitivity check (Section 4.9).

3.6. Analytic Procedures

3.6.1. Measurement Evaluation

Internal consistency was estimated with Cronbach's α alongside corrected item-total correlations and α -if-deleted diagnostics. Dimensionality was assessed by principal component analysis, with retention determined by Horn's parallel analysis (Horn, 1965) against the 95th percentile of eigenvalues from 500 simulated random datasets rather than by the Kaiser criterion alone. Parallel analysis was chosen deliberately, not as a fallback. It is the more conservative rule: unlike the Kaiser criterion, it holds the observed eigenvalue structure against a null distribution generated from data of the same rank and sample size, so a second component is retained only when it exceeds what random data of comparable dimensions would produce (Hayton et al., 2004). On that test, the seven indicators yielded a unidimensional solution, with one component accounting for 67.7% of item variance and the second observed eigenvalue (0.71) well below the simulated threshold. No confirmatory factor model was estimated. A separate confirmatory validation was not attempted because the available data did not permit an independent split-sample validation, and fitting a one-factor model to the same cases that produced the exploratory solution would not constitute independent confirmation. Average variance extracted and the heterotrait-monotrait ratio were not computed because both are used for discriminant comparisons between factors. With a single retained component, there is no such comparison to make.

Unidimensionality in cross-sectional ratings indicates that the indicators covary; it does not imply that they form a single substantively unified property. Because all seven are ratings of the respondent's own organization, part of their shared variance may reflect

a general evaluative disposition toward that organization. This interpretation is carried forward into the analysis of outcomes (Section 5.2).

3.6.2. Inferential Procedures

Associations between readiness and each outcome were estimated separately rather than against an aggregate index. Spearman's rank-order correlations were the primary estimator, reported with Fisher z -transformed 95% confidence intervals, with Pearson coefficients as a second-method check. Because five readiness–outcome associations are tested, Holm-adjusted p -values are reported as part of the sensitivity analysis. Adoption maturity was modeled using ordinal logistic regression with institution type and experience as covariates, and the proportional-odds assumption was tested with the Brant (1990) Wald test, for each predictor and globally. Subgroup comparisons used the Kruskal–Wallis and Mann–Whitney tests; categorical comparisons used the chi-square test with Cramér's V . All estimates reported in this revision were computed, or re-computed and verified against the harmonized dataset, in Python 3.11 (pandas 3.0, SciPy 1.17, statsmodels 0.15). Because many organizations were represented by a single respondent and others by several, multi-level models were not feasible; instead, the main estimates were re-estimated with a cluster bootstrap that resamples organizations (1000 resamples; unnamed respondents treated as their own clusters).

Pairwise deletion follows from the structure of the nonresponse rather than from convenience. Valid n declines across the instrument, from 330 on the early modules to 181 on the later ones (Section 4.9). Much of this decline is structural rather than behavioral: the 82 respondents to the earlier questionnaire version were not asked the outcome modules in their current form, and the primary readiness analyses use the two Arabic administrations of the current instrument ($n = 230$), which share wording, response format, and coding. The 24 English-stream responses are included in the descriptive analyses and a sensitivity check (Section 4.9). Within the current Arabic instrument, item nonresponse in the later modules was 3% for perceived Arabic-language output quality and training, and 20–21% for adoption maturity, return on investment, and expected future role. This residual nonresponse is patterned rather than diffuse. The downstream modules ask about training programs, return assessment, and the handling of Arabic output, each presupposing that the organization has AI arrangements to report on; journalists in legacy newsrooms running no active tools had nothing to report and passed them over. The missingness is, therefore, theoretically substantive rather than random, precisely the condition under which listwise deletion performs worst, since it would discard the low-adoption organizations wholesale and retain a sample tilted toward the better-equipped. Pairwise deletion does not correct this selection either: each estimate is computed on the subsample that answered the relevant items, and these subsamples become more selective in later modules. The valid n is, therefore, reported with every estimate, and three further checks bound the consequences: a comparison of later-module completers and non-completers on early-module variables, scenario bounds for the principal descriptive proportions, and listwise re-estimation (Section 4.9). None of these removes the bias; they indicate its likely direction and size.

3.7. Measurement Limitations

The instruments imposed structural constraints that bound the study's scope. Gender, age, nationality, and baseline AI awareness were not recorded, so those covariates cannot be examined. Open-ended responses were collected but, in the absence of a formal coding protocol and an intercoder reliability estimate, are not treated as analyzed data; where quoted below, they illustrate an interpretation carried by the quantitative estimates and

carry no independent evidentiary weight. Of the four open-ended questions, 177 respondents answered at least one; the questions quoted in Section 4.7.4 received 171 responses (examples of AI in the workflow), 150 (organizational initiatives), and 154 (suggestions). The first author selected quotations because they addressed how AI enters individual work and who governs its use, the themes relevant to RQ3; they are not presented as representative of the sample.

A second constraint concerns validation. The intended split-sample exploratory and confirmatory strategy could not be executed because the available data did not permit an independent split-sample validation. Structural validity, therefore, rests on internal consistency, item-level diagnostics, and parallel-analysis dimensionality assessment, and the measurement claims are bounded accordingly. The evidence supports the claim that the analysis depends on the assumption that the seven indicators behave as one dimension rather than four in these data. Formal content validity evidence (e.g., expert ratings of item relevance) was not collected.

3.8. Ethical Considerations

The study was approved by the Research Ethics Committee of Zayed University (protocol ZU25_074_F). Participation was voluntary. The information statement at the start of the questionnaire explained the study's purpose and assured respondents that their answers were confidential and would be used only for research purposes (in the current instrument, "research and development purposes"); consent was implied by completing the questionnaire after reading it. No names were requested. The organization-name item was used only to assess the clustering of respondents within organizations and is not reported. Quotations are reported without role, organization, or other identifying details, and were checked to ensure they do not allow a respondent or organization to be identified.

4. Results

All analyses use the harmonized dataset ($N = 336$). Item nonresponse varied across questionnaire modules, so the valid n is reported for each analysis.

4.1. Sample Description

The analytical sample comprised 336 media professionals across three streams ($n_1 = 267$, $n_2 = 45$, $n_3 = 24$); Table 1 reports experience, occupational role, and organization type by stream and in the pooled sample.

Nearly two-thirds (65.9%) reported more than ten years of experience, and 18.2% held leadership positions. Reporters, editors, correspondents, and photographers made up 41.2% of the sample; video producers and digital content creators, 9.7%; and 30.9% described other roles, including public relations, translation, presenting, and freelance media work. Respondents came from print or broadcast outlets (48.8%), digital-only platforms or news agencies (27.6%), government media offices (20.0%), and other organizations (3.6%). The sample is thus weighted toward experienced practitioners. Experienced journalists may judge AI output against more demanding editorial standards. They may view newsroom change through the lens of longer institutional memory, so levels of perceived quality and readiness should not be read as those of the workforce as a whole. Readiness did not differ across experience bands (Section 4.8.1), which limits, but does not remove, this concern.

Table 1. Respondent Characteristics for the Pooled Sample and Questionnaire Streams.

Characteristic	Pooled Sample (N = 336)	Stream 1 (n = 267)	Stream 2 (n = 45)	Stream 3 (n = 24)
Professional Experience (valid n = 328)				
<3 years	13.1% [9.9, 17.2]	9.7%	28.9%	20.8%
3–10 years	21.0% [17.0, 25.8]	17.4%	37.8%	29.2%
>10 years	65.9% [60.6, 70.8]	73.0%	33.3%	50.0%
Harmonized Occupational Role (valid n = 330)				
Reporter, editor, correspondent, or photographer	41.2% [36.0, 46.6]	45.2%	13.3%	50.0%
Video producer or digital content creator	9.7% [7.0, 13.4]	5.7%	31.1%	12.5%
Section head or platform/project manager (leadership)	18.2% [14.4, 22.7]	20.3%	8.9%	12.5%
Other roles	30.9% [26.2, 36.1]	28.7%	46.7%	25.0%
Harmonized Organization Type (valid n = 330)				
Print or broadcast	48.8% [43.4, 54.2]	52.5%	33.3%	37.5%
Digital-only platform or news agency	27.6% [23.0, 32.6]	21.1%	55.6%	45.8%
Government media office	20.0% [16.0, 24.6]	21.8%	11.1%	16.7%
Other	3.6% [2.1, 6.2]	4.6%	0.0%	0.0%

Note. Percentages use the valid *n* for each variable shown for each block; brackets are Wilson 95% confidence intervals. Roles, experience bands, and organization types follow the questionnaire’s own response options (Section 3.3); “leadership” is the section head or platform/project manager option used in Section 4.8.

4.2. Task-Level AI Use

A total of 218 respondents identified their primary use of AI (Table 2). Comparability testing indicated significant differences in task distributions across streams ($\chi^2(5) = 63.34$, $p < 0.001$, Cramér’s $V = 0.539$), so task distributions are presented by stream, with pooled totals reported for descriptive completeness only. The English stream is not shown because its version asked how often each task was performed rather than which task was primary, so its responses are not comparable.

Table 2. Primary AI Task by Questionnaire Stream.

Primary AI Task	Main Stream (n = 176)	Second Stream (n = 42)	Total (N = 218)
Information gathering and content generation	80 (45.5%)	1 (2.4%)	81 (37.2%)
Machine translation	30 (17.0%)	28 (66.7%)	58 (26.6%)
Fact-checking and deepfake detection	17 (9.7%)	11 (26.2%)	28 (12.8%)
Editing and text summarization	26 (14.8%)	1 (2.4%)	27 (12.4%)
Audience data analysis	17 (9.7%)	1 (2.4%)	18 (8.3%)
Drafting news reports and articles	6 (3.4%)	0 (0.0%)	6 (2.8%)

Note. $\chi^2(5) = 63.34$, $p < 0.001$, Cramér’s $V = 0.539$. Because distributions differed significantly between streams, stream-specific results are primary; pooled totals are descriptive. “Drafting news reports and articles” is not a questionnaire option; it was volunteered by respondents and is retained as a distinct category.

Task specialization differed substantially. In the main stream, information gathering and content generation predominated (45.5%, 95% CI [38.3, 52.8]); in the second, machine translation accounted for 66.7%. Fact-checking and deepfake detection accounted for 9.7% in the main stream and 26.2% in the second, while audience data analysis accounted for 9.7% and 2.4%, respectively.

4.3. Measurement Validation

4.3.1. Data Harmonization and Cross-Stream Comparability

Harmonization produced complete agreement between documented response options and the reconstructed coding scheme for organization type, training provision, adoption maturity and perceived Arabic-language output quality, the sole discrepancy being the undocumented “Good” category volunteered by respondents (Section 4.9).

Comparability testing showed that occupational role, experience, organization type, training provision, return on AI investment, and expected future role differed significantly across streams (Table 3), as did primary AI task (Table 2), whereas no significant differences were detected for adoption maturity, personal performance impact, perceived Arabic-language output quality, and skills needed (Table 3); the readiness composite did not differ between the two Arabic administrations (Section 4.9). Given the small streams, this is absence of evidence of a difference rather than evidence of equivalence (Section 3.4).

Table 3. Distributional Comparability Across Questionnaire Streams.

Item	χ^2	df	<i>p</i>	Cramér's V	Pooled for Primary Analysis
Occupational role	43.12	6	<0.001	0.256	No
Professional experience	30.54	4	<0.001	0.216	No
Organization type	35.93	4	<0.001	0.269	No
Training provision	43.10	8	<0.001	0.300	No
Return on AI investment	51.75	6	<0.001	0.367	No
Expected future role	10.48	4	0.033	0.241	No
Adoption maturity	8.36	4	0.079	0.214	Yes
Personal performance impact	9.04	5	0.107	0.223	Yes
Perceived Arabic-language output quality	10.53	10	0.395	0.146	Yes
Skills needed	1.79	3	0.618	0.090	Yes

Note. Expected cell counts below five exceeded 20% for several items. Monte Carlo *p*-values (5000 permutations) for the items entering the principal analyses were: occupational role < 0.001; professional experience < 0.001; organization type < 0.001; return on AI investment < 0.001; adoption maturity 0.092; personal performance impact 0.107; perceived Arabic-language output quality 0.372. No pooling decision changed (Supplementary Material S1). Organization type was tested on the current instrument's three categories (valid *n* = 249); valid *n* for every item is reported in S1. Variables designated “No” are reported both by questionnaire stream (Supplementary Material S2) and as pooled estimates.

4.3.2. Validation of the Institutional AI Readiness Scale

The seven-item scale demonstrated high internal consistency (Cronbach's α = 0.919, 95% CI [0.901, 0.934]; listwise *n* = 218). Corrected item–total correlations ranged from 0.605 to 0.809. Removing any indicator except actual adoption lowered α ; removing actual adoption, the indicator with the weakest item–total correlation, raised α marginally to 0.920 (Table 4).

Table 4. Institutional AI Readiness Indicators and Scale Properties.

Indicator	Indicator Type	<i>n</i>	M	SD	Corrected Item–Total <i>r</i>	α If Deleted
Actual adoption of AI in daily work	Implementation outcome	224	56.4	22.8	0.605	0.920
Staff media and digital qualification	Resource	222	53.7	22.9	0.781	0.904
Speed of converting ideas into applications	Process	276	56.5	26.4	0.809	0.900
Editorial–technical collaboration	Process	221	60.9	25.7	0.762	0.905
Use of audience engagement in development	Process	272	59.2	25.5	0.803	0.901
Availability of AI and misinformation guidance	Resource	270	54.7	29.0	0.791	0.902
Accessibility across audience groups	Normative commitment	223	55.9	28.1	0.703	0.912

Note. Indicators were rated on a 0–100 scale. Full-scale Cronbach's $\alpha = 0.919$, 95% CI [0.901, 0.934]; listwise $n = 218$. Indicator type follows the classification in Section 3.2.1. Component loadings, communalities, and the distribution of the number of indicators answered per respondent are reported in Supplementary Material S3.

The indicators co-varied strongly. Inter-item correlations ranged from 0.39 to 0.77 ($M = 0.62$), raw and standardized estimates differing negligibly ($\Delta\alpha = 0.0007$). Principal component analysis identified a dominant first component (eigenvalue = 4.74) explaining 67.7% of total variance; parallel analysis retained one factor, the second eigenvalue (0.71) falling below the simulated 95th-percentile threshold.

4.3.3. Evaluation of the Agreement Battery

Reconstruction of the agreement battery revealed uneven item completeness, yielding a listwise-complete subsample of 94 across the ten consistently populated items.

Overall internal consistency was acceptable ($\alpha = 0.789$), but the multidimensional structure specified in the original instrument was not supported: parallel analysis retained two factors explaining 27.3% and 23.4% of the variance, and the predefined subscales were unreliable (Organizational Culture $\alpha = 0.511$, Ethics and Governance $\alpha = 0.525$, Audience Orientation $\alpha = 0.421$). Only Perceived Media Performance reached acceptable consistency ($\alpha = 0.731$).

The battery was, therefore, retained for item-level descriptive analyses only and not modeled as a latent construct.

4.4. Descriptive Context: Depth of Institutional AI Integration

AI integration sits at intermediate maturity (Table 5). Nearly half of respondents (46.4%) reported AI used for simple or routine tasks and roughly a quarter (24.0%) regular use in selected workflows. Only 23 respondents (12.6%, 95% CI [8.5, 18.2]) described AI as essential or near-comprehensive.

Support structures were similarly thin. Of 240 respondents answering the training item, 65.4% (95% CI [59.2, 71.1]) reported training absent or insufficient and 25.4% (95% CI [20.3, 31.3]) judged it adequate.

Perceived returns followed the same pattern: of 192 valid responses, 65.1% (95% CI [58.1, 71.5]) reported returns that could not yet be assessed (45.8%) or had not materialized

(19.3%), and 12.5% (95% CI [8.5, 17.9]) reported a clear return. The composite readiness score averaged 56.6 (SD = 20.4), above the midpoint but well short of settled institutional capacity. Because these proportions rest on respondents who answered later modules, they are best read with the scenario bounds reported in Section 4.9.

Table 5. Institutional AI Adoption Maturity ($n = 183$).

Stage	Description	n	%	95% CI
5	Near-total reliance	3	1.6	[0.6, 4.7]
4	Essential element of the workflow	20	10.9	[7.2, 16.3]
3	Regular use in some editorial or digital processes	44	24.0	[18.4, 30.7]
2	Partial use in simple or routine tasks	85	46.4	[39.4, 53.7]
1	Limited or experimental use	31	16.9	[12.2, 23.0]

Note. Percentages and 95% confidence intervals are calculated using the valid sample size for the adoption maturity item ($n = 183$).

4.5. RQ1: Reliability and Dimensionality of Institutional Readiness

The readiness indicators showed high internal consistency (Section 4.3.2). What reliability cannot settle is whether the conceptual domains behind the item pool correspond to anything separable in the data.

Dimensionality told a different story. A single component dominated and parallel analysis retained one factor, the second observed eigenvalue (0.71) falling below both the Kaiser criterion and the simulated threshold.

The four conceptual domains were, therefore, not recovered as separable dimensions. In these cross-sectional data, the indicators co-occur as a single dimension, and the composite is used below as a single scalar rather than as a diagnostic profile from which particular weaknesses can be read off. This does not establish that readiness is one substantively unified organizational property: the indicators span resources, processes, a normative commitment, and one implementation outcome (Table 4), and their co-variation may partly reflect respondents' overall evaluation of their organization.

4.6. RQ2: Institutional AI Readiness and Organizational Outcomes

4.6.1. Bivariate Associations Between Institutional AI Readiness and Outcome Measures

Bivariate analysis revealed a clear split between internal organizational outcomes and perceived Arabic-language output quality (Table 6; Figure 1).

Table 6. Associations Between Institutional AI Readiness, Internal Organizational Outcomes, and Perceived Arabic-Language Output Quality.

Outcome Measure	Spearman's ρ	95% CI	n	p	Holm-Adjusted p	Pearson r
Perceived AI return on investment (ROI)	0.431	[0.253, 0.580]	97	<0.001	<0.001	0.433
Training provision	0.429	[0.306, 0.537]	193	<0.001	<0.001	0.432
Adoption maturity	0.415	[0.285, 0.529]	179	<0.001	<0.001	0.425
Personal performance impact	0.226	[0.069, 0.373]	150	0.005	0.011	0.223
Perceived Arabic-language output quality	0.097	[−0.038, 0.229]	212	0.159	0.159	0.103

Note. Spearman's ρ reported with Fisher z -transformed 95% confidence intervals; Pearson r provided as a second-method check. Holm adjustment across the five tests. Estimates with the six-indicator composite and adjusted for questionnaire stream are reported in Section 4.9. Because the correlations are computed on different, overlapping subsamples, differences between each organizational correlation and the correlation with perceived Arabic-language output quality were tested by bootstrap (2000 resamples of respondents).

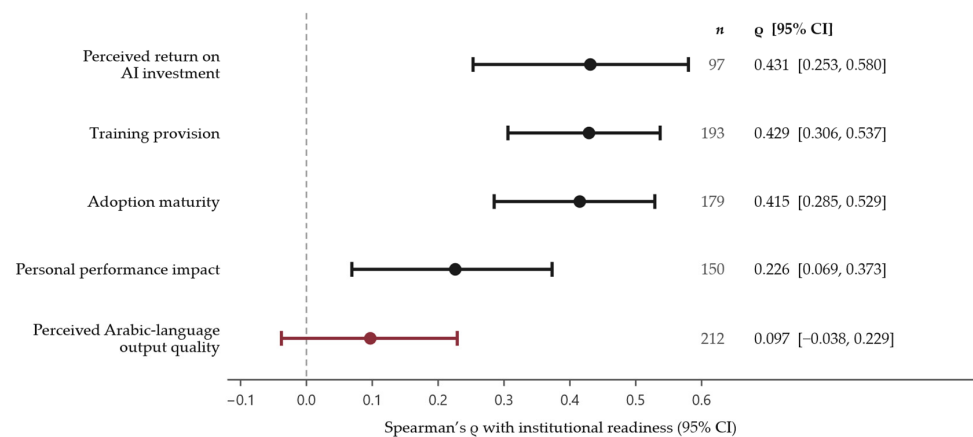


Figure 1. Associations between institutional readiness and five outcome measures, with 95% confidence intervals. Intervals excluding zero are shown in black; the interval spanning zero is shown in red. Estimates correspond to Table 6; the Arabic outcome is practitioners' perceived output quality.

Three internal organizational outcomes were moderately and highly significantly associated with readiness: perceived return on AI investment ($\rho = 0.431$, 95% CI [0.253, 0.580], $n = 97$), training provision ($\rho = 0.429$, 95% CI [0.306, 0.537], $n = 193$), and adoption maturity ($\rho = 0.415$, 95% CI [0.285, 0.529], $n = 179$), all $p < 0.001$. Perceived personal performance effect was weaker but significant ($\rho = 0.226$, 95% CI [0.069, 0.373], $n = 150$, $p = 0.005$), and all four associations remained significant after Holm adjustment. The contrast is itself statistically supported: bootstrap 95% intervals for the difference from the correlation with perceived Arabic-language output quality excluded zero for return on investment ($\Delta\rho = 0.333$ [0.098, 0.546]), training provision ($\Delta\rho = 0.332$ [0.134, 0.519]), and adoption maturity ($\Delta\rho = 0.317$ [0.126, 0.494]), but not for personal performance ($\Delta\rho = 0.129$ [−0.054, 0.303]).

Perceived Arabic-language output quality showed no significant association with readiness ($\rho = 0.097$, 95% CI [−0.038, 0.229], $n = 212$, $p = 0.159$). The interval is the substantive result rather than the point estimate: it excludes moderate and larger associations while remaining compatible with a small one. The results therefore suggest a distinction between organizational outcomes and practitioners' assessments of AI output.

4.6.2. Ordinal Logistic Regression Model of Adoption Maturity

Readiness was the sole significant predictor (Table 7). A one-standard-deviation increase was associated with 2.59 times higher odds of occupying a higher adoption maturity category (OR = 2.587, 95% CI [1.879, 3.561], $p < 0.001$); neither institution type nor experience contributed independently.

Table 7. Ordinal Logistic Regression Predicting AI Adoption Maturity ($n = 179$).

Predictor Term	Odds Ratio (OR)	95% CI	Parameter (b)	SE	p
Institutional Readiness (per SD)	2.587	[1.879, 3.561]	0.950	0.163	<0.001
Institution: Government Media	1.020	[0.429, 2.424]	0.019	0.442	0.965
Institution: Print/Broadcast	0.941	[0.453, 1.957]	−0.061	0.373	0.871
Experience: <3 Years	0.657	[0.260, 1.660]	−0.420	0.473	0.374
Experience: 3–10 Years	0.778	[0.395, 1.532]	−0.251	0.346	0.467

Note. Model LR $\chi^2(5) = 41.46$, $p < 0.001$; McFadden pseudo- $R^2 = 0.092$. Reference categories: Digital-only platform/News agency (Institution); >10 years (Experience). The model uses respondents to the two Arabic administrations of the current instrument, whose institution and experience categories are shown (Section 3.3). Adoption stages 4 and 5 were merged for estimation because stage 5 had three respondents; thresholds (logit scale) were −2.01, 0.54, and 2.15. Models without the adoption indicator and with questionnaire stream are reported in Section 4.9.

The proportional-odds assumption was not rejected, either globally (Brant test $\chi^2(10) = 8.63$, $p = 0.568$) or for the readiness coefficient ($\chi^2(2) = 1.08$, $p = 0.583$); no other predictor approached significance (all $p > 0.50$).

4.7. RQ3: Institutional AI Readiness and Practitioners' Perceived Quality of Arabic-Language AI Output

4.7.1. Descriptive Evaluation of Perceived Content Quality

Most respondents rated Arabic AI-generated content average (58.9%, 95% CI [52.7, 64.9]; Table 8); 27.6% (95% CI [22.4, 33.5]) rated it excellent or very excellent (30.9% if the off-scale "Good" responses are included) and 10.2% (95% CI [7.0, 14.6]) poor or worse.

Table 8. Practitioners' Perceived Quality of Arabic-Language AI Output ($n = 246$).

Rating Category	<i>n</i>	%	95% CI
Very excellent	11	4.5	[2.5, 7.8]
Excellent	57	23.2	[18.3, 28.8]
Good (off-scale response)	8	3.3	[1.7, 6.3]
Average	145	58.9	[52.7, 64.9]
Poor	19	7.7	[5.0, 11.7]
Very poor	6	2.4	[1.1, 5.2]

Note. "Good" was volunteered by respondents and is not a documented scale option; it is retained as a distinct category (see Section 4.9). All eight "Good" ratings came from the main stream and were among 20 responses entered on a single day, consistent with the transcription of face-to-face responses in which respondents offered a rating between the documented options. To that extent the item was not administered as a strictly closed scale.

4.7.2. Institutional Readiness and Perceived Quality Evaluations

Institutional AI Readiness was not significantly associated with perceived Arabic-language output quality ($\rho = 0.097$, 95% CI $[-0.038, 0.229]$, $n = 212$, $p = 0.159$); with this sample the study had 80% power to detect $\rho \geq 0.194$. None of the seven individual indicators was significantly associated with perceived output quality either (ρ range 0.007 to 0.112, all $p > 0.10$, $n = 211$ –215), so the null does not reflect aggregation across heterogeneous indicators.

4.7.3. Organizational and Individual Correlates of Perceived Content Quality

Although readiness was not significantly associated with perceived Arabic-language output quality, practitioners' evaluations were positively associated with perceived personal performance effect ($\rho = 0.231$, 95% CI $[0.074, 0.377]$, $n = 151$, $p = 0.004$). No association was observed with adoption maturity ($\rho = 0.107$, $p = 0.154$), training provision ($\rho = 0.013$, $p = 0.845$), or perceived return on investment ($\rho = 0.106$, $p = 0.146$). Perceived Arabic-language output quality was thus associated with practitioners' perceived personal performance effect but not significantly with readiness or with the other organizational outcomes.

4.7.4. Practitioner Accounts of Arabic-Language Work

Open-ended responses, collected but not formally coded (Section 3.7), read consistently with the quantitative pattern and are quoted for illustration only. Asked for specific examples of how AI enters their work, respondents described a practice they had bounded for themselves. One drew the line at authorship: "I use it for rewrites, editing, proofreading my content, editing content for social media (never drafting new content at any point)... sometimes translation." Another, asked for suggestions, placed the safeguard after the tool rather than around it: "We can use AI but always do the necessary edits and checks before publication." Neither is describing a policy. Both are describing a rule they apply themselves.

Asked separately to name AI projects or initiatives at their organization, one respondent answered by denying the premise of the question: «معظم المبادرات اجتهادات شخصية وليست مؤسسية» (most initiatives are personal exertions of judgement, not institutional ones). The term reached for here, *ijtihad*, carries the sense of independent reasoned effort undertaken where no settled rule applies, and it names the position the other accounts describe occupying. A practitioner working in Arabic-language production was more specific about the isolation involved: «كفرد كوني بعيدة عن المطبخ التحريري استخدمه في مراجعة الأخبار وتحريرها» (individual, since I am far from the editorial kitchen; I use it to review and edit the news). Verifying Arabic output, catching dialectal slippage, judging which tasks a model can be trusted with: in these accounts, this is labour performed at the desk, by individuals, on their own authority.

The accounts sit well with the estimates. Readiness was not significantly associated with perceived Arabic-language output quality ($\rho = 0.097$) while being associated with each organizational outcome, and rated quality tracked practitioners' own sense of their performance ($\rho = 0.231$). Where the burden of linguistic verification falls on individual prudence, better organization may have little to substitute for it. These quotations cannot establish that reading; they offer illustrative material consistent with one possible interpretation of the correlational null, and they come from a small number of respondents.

4.8. RQ4 (Exploratory): Variation by Institution Type, Leadership Position, and Professional Experience

4.8.1. Institutional AI Readiness Across Organizational Groups

Institutional AI Readiness differed significantly across organization types (Kruskal–Wallis $H = 6.90$, $df = 2$, $p = 0.031$): government media scored highest ($M = 62.5$, $n = 50$), followed by print/broadcast ($M = 55.3$, $n = 119$) and digital-only platforms and news agencies ($M = 54.0$, $n = 52$). These groups follow the main questionnaire version's organization-type item (Section 3.3); the comparison with the six-indicator composite is reported in Section 4.9.

Readiness did not differ by leadership position (leaders $M = 56.9$, $n = 38$; non-leaders $M = 56.6$, $n = 183$; Mann–Whitney U , $p = 0.669$) or across experience bands ($H = 0.40$, $df = 2$, $p = 0.808$). The 60 respondents in leadership roles are distributed unevenly across items (Table 9).

Table 9. Leadership Differences Across AI-Related Organizational Outcomes.

Outcome/Variable	Leader <i>n</i>	Non-Leader <i>n</i>	χ^2	df	<i>p</i>	Cramér's <i>V</i>
Adoption maturity	30	153	1.04	4	0.904	0.075
Perceived AI return on investment	32	160	0.80	3	0.851	0.064
Training provision	40	200	5.63	4	0.229	0.153
Institution type	41	208	3.46	2	0.178	0.118
Personal performance impact	29	153	7.83	5	0.166	0.207
Perceived Arabic-language output quality	41	205	12.51	5	0.028	0.226
Expected future role	29	152	9.96	4	0.041	0.235
Experience	60	268	19.47	2	<0.001	0.244

Note. Comparisons use the chi-square test of independence. Expected future role was among the items failing the cross-stream comparability criterion in Table 3 ($\chi^2(4) = 10.48$, $p = 0.033$, Cramér's $V = 0.241$) and is pooled here. A robustness check confined to the main questionnaire stream, the largest of the three, returned the same result in direction and significance ($\chi^2 = 9.42$, $p = 0.044$). Adoption maturity, personal performance impact, and perceived Arabic-language output quality met the criterion; training provision, perceived return on investment, institution type, and experience did not (Table 3) and are pooled here, so their comparisons should be read accordingly. These comparisons are exploratory and unadjusted for multiplicity.

4.8.2. Leadership Differences Across Organizational Outcomes

Leadership status was not significantly associated with adoption maturity, perceived return on AI investment, training provision, institution type, and perceived personal performance effect (all $p > 0.05$).

Three differences did emerge. Leaders were substantially more experienced (90.0% with more than ten years, against 60.4%; $\chi^2 = 19.47$, $p < 0.001$, $V = 0.244$); they judged Arabic AI-generated content more critically (24.4% poor or very poor, against 7.3%; $\chi^2 = 12.51$, $p = 0.028$, $V = 0.226$); and they more often expected AI to change their role fundamentally (37.9% vs. 19.7%), while non-leaders more often reported uncertainty (20.4% vs. 3.4%; $\chi^2 = 9.96$, $p = 0.041$, $V = 0.235$). Because expected future role differed across versions (Table 3), this comparison was re-estimated within the main stream alone; direction and significance held ($\chi^2 = 9.42$, $p = 0.044$).

In these exploratory comparisons, readiness differed by organization type but not significantly by leadership position or experience. What distinguished leadership was assessment of AI output and experience, not organizational readiness or adoption depth.

4.9. Sensitivity and Robustness Analyses

Four pre-specified checks tested robustness: replication using the main questionnaire stream alone (valid $n = 155$ – 180 per estimate); alternative treatment of the undocumented “Good” category; listwise rather than pairwise deletion ($n = 76$); and a strict seven-of-seven composite inclusion rule. Full detail appears in the Supplementary Materials (S4–S7).

None altered a conclusion. Single-stream estimates fell within two percentage points of pooled values on four of five indicators; off-scale handling moved the share rated excellent or better between 27.6% and 30.9%; under listwise deletion, the readiness–outcome associations retained their direction and significance, including the Arabic-language null ($\rho = 0.100$, $p = 0.388$). Because the complete-case subsample is likely more selective than the pairwise subsamples, this shows only that the pattern does not depend on the deletion rule, not that the estimates are unbiased; and the strict and primary composites correlated at $r = 1.0000$, the adoption-maturity odds ratio shifting only from 2.587 to 2.585.

Case loss and selection. Table 10 reports, for each module, how many respondents were asked the item and how many answered, and bounds the principal descriptive proportions under two scenarios for the non-answers. If every respondent who skipped the adoption item worked at the lowest stage, the share reporting AI as essential or near-comprehensive would fall from 12.6% to 10.0%. Among the 230 respondents to the current Arabic instrument, those who answered the adoption item ($n = 183$) did not differ from those who skipped it ($n = 47$) in organization type ($\chi^2(2) = 1.95$, $p = 0.377$), readiness ($M = 57.2$ vs. 54.2 , Mann–Whitney $p = 0.469$), or perceived Arabic-language output quality ($p = 0.252$). They did differ by stream (skipping was more common in the second Arabic administration; $\chi^2(1) = 9.07$, $p = 0.003$) and by experience (more common among those with fewer than 3 years; $\chi^2(2) = 8.24$, $p = 0.016$). Weighting completers by the inverse of their predicted response probability from stream, organization type, and experience left the readiness–adoption association essentially unchanged (weighted $\rho = 0.408$ vs. 0.415).

Clustering and the English stream. Resampling organizations rather than respondents changed the intervals only slightly and changed no conclusion: adoption maturity $\rho = 0.415$ [0.235, 0.555]; training $\rho = 0.429$ [0.299, 0.545]; return on investment $\rho = 0.431$ [0.249, 0.587]; personal performance $\rho = 0.226$ [0.060, 0.341]; perceived Arabic-language output quality $\rho = 0.097$ [−0.056, 0.238]; odds ratio 2.59 [1.74, 4.00]. Adding the English stream’s readiness responses (composite $n = 245$) also left the pattern unchanged (e.g., training $\rho = 0.435$; perceived Arabic-language output quality $\rho = 0.110$, $p = 0.092$; Supplementary Material S9).

Table 10. Items Asked, Valid Responses, and Scenario Bounds for Descriptive Proportions.

Module/Estimate	Asked of (<i>n</i>)	Valid <i>n</i>	Observed %	Bound If All Non-Answers Fall Outside the Category	Bound If All Non-Answers Fall Inside the Category
Occupational role	336	330	-	-	-
Readiness composite (≥ 5 of 7 indicators)	230	221	-	-	-
Primary AI task	230	218	-	-	-
Perceived Arabic-language output quality: average or worse	254	246	69.1	66.9	70.1
Training provision: absent or insufficient	254	240	65.4	61.8	67.3
Perceived return: clear return	254	192	12.5	9.4	33.9
Adoption maturity: essential or near-total (stages 4–5)	230	183	12.6	10.0	30.4
Expected future role	230	181	-	-	-

Note. “Asked of” counts respondents whose questionnaire version contained the item and whose responses entered that analysis: the 230 respondents to the two Arabic administrations of the current instrument, plus the 24 English-stream respondents, where their answers are pooled. Bounds re-express each proportion over the respondents asked. If skipping reflects the absence of AI arrangements (Section 3.6.2), the relevant bounds are the first for adoption and clear return, and the second for absent or insufficient training.

Readiness without the adoption indicator and stream-adjusted estimates. Table 11 re-estimates every readiness–outcome association with a six-indicator composite that omits actual adoption, and re-estimates the ordinal model with harmonized categories and questionnaire stream as a covariate. Excluding the adoption indicator left the scale’s properties intact ($\alpha = 0.920$; one component explaining 71.8% of variance) and reduced the odds ratio for adoption maturity from 2.59 to 2.38 (95% CI [1.74, 3.25]); the correlation with adoption maturity fell from 0.415 to 0.390. All four organizational associations remained significant after Holm adjustment, and the association with perceived Arabic-language output quality remained non-significant ($p = 0.095$, $p = 0.166$). Adjusting for questionnaire stream strengthened, rather than weakened, the readiness estimates (odds ratio 2.71 with seven indicators; 2.48 with six). Adoption maturity was lower in the second Arabic administration at a given level of readiness (OR = 0.19 [0.07, 0.53]), but readiness itself did not differ between the two administrations (Mann–Whitney $p = 0.880$). Full model output is provided in Supplementary Material S9.

Table 11. Re-estimation Without the Adoption Indicator and With Stream Adjustment.

Estimate	Seven Indicators (as Reported)	Six Indicators, Without Actual Adoption	Seven Indicators, Stream-Adjusted	Six Indicators, Stream-Adjusted
Cronbach’s α [95% CI] (listwise <i>n</i>)	0.919 [0.901, 0.934] (218)	0.920 [0.902, 0.935] (218)	-	-
First eigenvalue/% variance/second eigenvalue	4.74/67.7%/0.71	4.31/71.8%/0.49	-	-
ρ with perceived ROI (<i>n</i> = 97)	0.431 ***	0.434 ***	0.458 ***	0.463 ***

Table 11. Cont.

Estimate	Seven Indicators (as Reported)	Six Indicators, Without Actual Adoption	Seven Indicators, Stream-Adjusted	Six Indicators, Stream-Adjusted
ρ with training provision ($n = 193$)	0.429 ***	0.423 ***	0.429 ***	0.423 ***
ρ with adoption maturity ($n = 179$)	0.415 ***	0.390 ***	0.429 ***	0.404 ***
ρ with personal performance impact ($n = 150$)	0.226 **	0.208 *	0.230 **	0.211 *
ρ with perceived Arabic-language output quality ($n = 212$)	0.097	0.095	0.097	0.095
OR for adoption maturity, per SD [95% CI]	2.587 [1.879, 3.561]	2.380 [1.742, 3.250]	2.711 [1.955, 3.760]	2.484 [1.808, 3.413]
Brant test, global χ^2 (df), p	8.63 (10), 0.568	8.20 (10), 0.609	9.76 (12), 0.637	9.50 (12), 0.660
Kruskal–Wallis H by organization type (p)	6.90 (0.031)	6.22 (0.045)	-	-

Note. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; all organizational associations remain significant after Holm adjustment. Stream-adjusted correlations are partial Spearman correlations controlling for questionnaire stream (main vs. second Arabic administration); stream-adjusted ordinal models add stream to the institution and experience covariates of Table 7. The six-indicator composite requires at least four of six indicators ($n = 221$) and correlates 0.99 with the seven-indicator composite. α confidence intervals follow Feldt's method.

4.10. Exploratory Analyses

These analyses were not pre-specified, were not adjusted for multiple comparisons, and are interpreted as exploratory.

Examined by organization type, the readiness–perceived Arabic-language output quality association was positive and significant only within digital-only platforms and news agencies ($\rho = 0.284$, $n = 51$, $p = 0.043$), near zero in government media ($\rho = -0.016$, $n = 48$) and in print and broadcast organizations ($\rho = 0.064$, $n = 113$), and did not survive Bonferroni adjustment across the three organization types ($\alpha = 0.05/3 \approx 0.017$).

Exploratory analysis of the reconstructed agreement battery suggested a two-factor solution but rested on a listwise-complete subsample of 94; no conclusions are drawn from it.

4.11. Summary of Findings

Institutional AI Readiness was associated with internal organizational outcomes (adoption depth, training provision, and perceived return on AI investment) and with perceived professional performance, but not with practitioners' perceived quality of Arabic-language AI output. In exploratory comparisons, readiness differed by institution type but not significantly by hierarchical position or professional experience.

5. Discussion

The discussion separates four kinds of statement: what was observed, what the observations are consistent with, which mechanisms remain unverified, and which causal claims require further research.

5.1. Principal Findings

In these data, perceived institutional readiness for AI was treated as a single measurable dimension rather than a set of separable capacities. Seven indicators, from technology use and staff qualifications through implementation speed, editorial–technical collaboration, governance provision, and accessibility, formed a highly reliable scale, with one com-

ponent accounting for 67.7% of the variance. The indicators co-varied closely; whether they would also change together under intervention is a separate question the design cannot answer.

That dimension is associated with how deeply AI is reported to be embedded, but reported AI integration was concentrated at the partial-use stage. A one-standard-deviation increase in readiness was associated with 2.59 times higher odds of a higher adoption stage. However, only 12.6% of those who answered the adoption item ($n = 183$) described AI as essential or near-comprehensive. Among respondents answering the relevant items, about two-thirds reported absent or insufficient training (65.4%, $n = 240$) and about two-thirds reported that return on investment was absent or could not yet be assessed (65.1%, $n = 192$).

Readiness was associated with the organizational outcomes. It was not associated with practitioners' perceived Arabic-language output quality to any comparable degree: its association with practitioners' ratings was not distinguishable from zero, with an interval excluding moderate and larger effects while permitting a small one.

Comparison with prior research. The training deficit reported here matches findings elsewhere. The global JournalismAI survey identified financial constraints and a lack of resources for qualified personnel and upskilling programs as key barriers to AI adoption (Beckett & Yaseen, 2023). In the United Kingdom, Thurman et al. (2025) found that AI use and training provision reflect how well-resourced a news organization is. In the UAE, professionals interviewed by Hassouni and Mellor (2025) reported no clear guidelines or training plans and relied on self-directed learning. The positive association between readiness and training provision ($\rho = 0.429$) is consistent with these accounts. The present data depart from them in one respect. Hassouni and Mellor (2025) expressed concern about the quality of AI-generated Arabic, and practitioners here were similarly reserved (average or worse in 69.1% of cases); yet that judgment was not significantly associated with the rated readiness of their organizations. Earlier newsroom studies documented organizational conditions shaping uptake (Thurman et al., 2017; Milosavljević & Vobič, 2019). The present results suggest that this organizational account holds for implementation outcomes but not for perceived output quality.

5.2. Theoretical Implications

What was observed? Within these data, institutional readiness behaved as a unidimensional measure and was associated with four organizational outcomes. Three associations are moderate ($\rho = 0.415$ – 0.431 , all $p < 0.001$); the fourth, perceived personal performance effect, is weaker ($\rho = 0.226$). The association with perceived Arabic-language output quality is bounded below moderate magnitude.

What the evidence is consistent with. The pattern accords with sociotechnical accounts in which outcomes depend on the joint state of technical and social subsystems (Trist & Bamforth, 1951), and with the view that technology and structure are mutually constituting (Orlikowski, 1992; Leonardi, 2011). An organization can restructure workflows and retrain staff, whereas it cannot directly alter the underlying Arabic-language capabilities of the models it uses. The observed asymmetry is consistent with this distinction. Generative systems press on norms of authorship and professional identity that no newsroom governs alone (Lewis et al., 2025).

What remains unverified. A central alternative concerns the composite itself. Its reliability is high across substantively distinct indicators, suggesting that it partly captures a general evaluative disposition toward the organization rather than four independent capacities. That reading fits the outcome pattern: readiness was associated with other organizational self-assessments but not with an assessment of something external to the organization. Common-method variance (Podsakoff et al., 2003) remains a viable explanation for

the positive associations, though it does not account for the null. The conceptual overlap between the adoption indicator and adoption maturity is a second threat, which the six-indicator re-estimation addresses (Table 11). Mechanisms linking readiness to outcomes were not observed.

What requires further research? The associations are cross-sectional; whether readiness precedes adoption, follows it, or both reflects a common cause cannot be settled here (Section 5.6).

5.3. Boundary Conditions

The most consequential result is not that readiness is associated with adoption, but that its reach has an identifiable edge falling along a theoretically meaningful line. Institutional readiness appears to have a bounded explanatory domain: in these data, it is associated with organizational capabilities and implementation outcomes more consistently than with practitioners' perceived output quality. This is not a claim that organizational factors are irrelevant to how well AI performs, but rather that the class of outcomes determines how strongly readiness is associated with those outcomes. Readiness constructs may be better suited to explaining whether a technology becomes embedded than to explaining perceived performance in a domain-specific task.

The distinction is visible in which outcomes moved. Adoption depth, training provision, and perceived return on AI investment concern organizational arrangements: things a newsroom decides, funds, and structures. Perceived Arabic-language output quality may partly depend on model capability, Arabic training data, and the handling of morphology and diglossia, none of which a newsroom controls by reorganizing itself. Because output quality was perceived rather than measured, this link remains an interpretation rather than an observation.

If replicated, the implication would extend beyond this sample. Studies treating readiness as a global explanatory variable may overstate its reach by aggregating outcome classes that respond differently to it; where implementation and perceived output outcomes are pooled into a single dependent construct, a boundary of this kind is averaged away rather than detected. The organizational half of the pattern is not peculiar to this sample: a Reuters Institute survey of UK journalists likewise reports that AI use, training provision, and dependence on third-party tools track how well resourced a news organization is (Thurman et al., 2025).

Two qualifications accompany this reading. First, the null is bounded, not absolute: the interval extends to $\rho = 0.229$, ruling out a moderate or large association with perceived Arabic-language output quality but not a small one. The defensible claim is limited explanatory reach, not irrelevance.

Second, the null is not uniform across contexts. In digital-only platforms and news agencies, readiness was associated with perceived Arabic-language output quality ($\rho = 0.284$, $n = 51$), whereas the association was near zero in traditional outlets. The subgroup is small and would not survive multiplicity correction, so it is reported as a hypothesis rather than a finding.

A third boundary concerns hierarchy. Leadership status was not significantly associated with readiness, adoption maturity, perceived return on investment, or training provision. Leaders were more experienced, rated Arabic-language output lower, and were markedly less likely to describe their trajectory as unclear. That readiness does not vary by position is worth registering against work reporting that professional role conditions whether generative tools are adopted at all (Møller et al., 2025): role may shape individual uptake without shaping how the organization is appraised. The absence of hierarchical differences should be read with its power limits in mind.

A fourth concerns experience. Two-thirds of respondents (65.9%, $n = 328$) had more than ten years in the profession, and leaders, who were almost all in that band, rated Arabic output more critically than non-leaders. Experienced practitioners may bring more demanding editorial standards to AI output, so the low perceived-quality ratings may partly reflect the evaluators as well as the systems. Readiness did not vary by experience, but a less experienced sample might rate output quality more favorably.

5.4. Confidence in the Reported Findings

The dataset arrived as three collection streams across two questionnaire versions, one of which carried numeric codes without value labels. The numeric stream was decoded by recovering the codebook from the original instrument and verifying it against code-to-label co-occurrence in a stream where both appeared, so the mapping rests on documentary and observed evidence rather than inference. Harmonization raised the item-level bases by 23 to 45 respondents without adding an imputed value.

5.5. Practical Implications

For newsroom leadership, readiness was associated with the depth of adoption, training provision, and perceived return on AI investment. The present cross-sectional measurement structure does not support interpreting the four proposed domains as empirically independent subscales. Intervention studies would be required to determine whether individual organizational capacities can be modified independently and whether improvements in one carry over to the others.

For journalists, perceived Arabic-language output quality was associated with perceived personal performance effect but not significantly with readiness. One interpretation, which these data cannot verify, is that practitioners working in Arabic treat output quality as a property of the tools rather than of the newsroom, so that task selection and verification at the individual level carry weight that organizational provision does not substitute for. One respondent described the position in those terms, as personal *ijtihad* rather than institutional practice (Section 4.7.4).

For those building Arabic-language systems and the researchers evaluating them, 69.1% of practitioners who rated it ($n = 246$) judged Arabic output as average or worse, and this perceived quality rating was not associated with greater institutional readiness. The evaluation literature independently reports that Arabic performance still trails English (Almatham et al., 2025); what these data add is that practitioners in production also perceive limitations, and that the gap between computational evaluation and working assessment is itself a target for investigation.

5.6. Limitations

Measurement: Only Institutional AI Readiness could be measured with a multi-item scale; the agreement battery's subscales had unacceptable reliability (Section 4.3.3). Adoption depth, training provision, perceived return on AI investment, perceived performance effect, and perceived Arabic-language output quality were each a single item. These are concrete organizational judgments rather than latent constructs, so single indicators are defensible; nonetheless, this precludes reliability estimation and attenuates the reported associations. Perceived Arabic-language output quality was a single general perception that did not specify tool, task, variety of Arabic, or criterion; direct technical performance was not observed. The first author translated the English version without independent back-translation. The readiness composite includes one indicator of current AI use, which overlaps with adoption maturity (addressed in Table 11), and its indicators mix resources, processes, and outcomes. All measures are self-reports obtained from a single instrument (Podsakoff et al., 2003).

Design: The design is cross-sectional and cannot establish temporal order. The readiness–outcome analyses rest on the 230 respondents to the current Arabic instrument, because the earlier version did not include the outcome modules in their current form. Within that group, item nonresponse in the later modules ranged from 20% to 21% and is likely not random: respondents in organizations without AI arrangements plausibly skipped items presupposing AI, and skipping was more common in the second administration and among less experienced respondents. Neither pairwise nor listwise deletion corrects this. Its likely effect is to bias upward adoption, training, and return estimates; completers and non-completers did not differ in readiness, and response weighting left the adoption association unchanged (Section 4.9). Reported adoption levels should therefore be read as upper bounds (Table 10), and leadership comparisons are underpowered. Pooling three streams assumes comparability that small-stream tests cannot confirm (Section 3.4). Respondents were nested in organizations; multilevel estimation was not feasible, although an organization-level bootstrap left the conclusions unchanged.

Scope: The findings describe media professionals in the United Arab Emirates and should not be generalized to Arabic-language media systems with different levels of state investment, infrastructure, or market structure. Recruitment was purposive, and the sample was weighted toward experienced practitioners (65.9% with more than ten years), so descriptive proportions characterize these respondents rather than the sector.

5.7. Future Research and Synthesis

Does readiness precede adoption? The association is observed; its direction is not. A longitudinal design that tracks newsrooms through the introduction of governance frameworks or training would help separate the competing accounts.

Two further questions concern who and where. The digital-native pattern rests on 51 respondents and would not survive multiplicity correction, so testing it requires a design stratified by organizational type. Leaders rated Arabic-language output poorly at more than three times the non-leader rate, a difference that could equally be explained by greater exposure to failure cases and differing evaluative standards.

Does practitioner assessment correspond to systematic evaluation? Pairing practitioner ratings with systematic evaluation of model output and testing whether the composite predicts behavioral measures, such as system logs or correction rates, would establish whether the boundary lies with the technology or with its perception.

Taken together, the findings point to three levels of constraint that are not interchangeable. Organizational constraints are the ones institutions can act on: readiness was associated with adoption depth, training provision, and perceived return on AI investment. Individual constraints operate at the practitioner's encounter with the tools: perceived Arabic-language output quality was associated with perceived personal performance effect ($p = 0.231$) but not significantly with readiness. Constraints located in the technology itself were not measured directly; practitioners' perceptions of output quality were the outcome least associated with organizational arrangements.

Interventions aimed at one level should not automatically be expected to relieve constraints at another. On this evidence, readiness was associated with organizational-level outcomes; whether governance or training reform affects perceived output quality remains to be tested.

6. Conclusions

This study asked whether institutional readiness can be measured as a coherent property and how far that property extends across the outcomes that readiness frameworks explain, with the reported depth of AI integration as context. Among respondents who an-

swered the adoption item ($n = 183$), reported AI integration was concentrated at the partial-use stage (46.4%), and 12.6% described AI as essential or near-comprehensive. Among those answering the relevant items, about two-thirds reported training as absent or insufficient ($n = 240$) and about two-thirds reported returns as absent or not yet assessable ($n = 192$). Readiness was measured as a single, reliable dimension associated with adoption depth, training provision, perceived return on AI investment, and perceived personal effect, but not significantly with practitioners' ratings of Arabic-language AI output.

The contribution lies in the contrast rather than in either result alone. In these data, institutional readiness appears to have a bounded explanatory domain: it is more consistently associated with organizational capabilities and implementation outcomes than with practitioners' perceived output quality. Studies that aggregate heterogeneous outcomes into a composite index may overstate its reach.

These conclusions rest on practitioners' perceptions gathered from a purposive, experienced sample, with outcome items answered by a subset of respondents. Perceived Arabic-language output quality was a single general rating, and direct technical performance was not observed. The readiness composite includes one indicator that overlaps with adoption; however, the sensitivity analysis excluding this indicator produced comparable results (Section 4.9; Table 11).

Interventions should be matched to the level at which a constraint operates. Governance provisions, staff qualifications, cross-team coordination, and implementation capacity are arrangements a newsroom can change, and these are the conditions with which readiness was associated. Perceived Arabic-language output quality was not significantly associated with institutional readiness, so the present findings do not support assuming that greater perceived institutional readiness alone will improve it.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/journalmedia7040212/s1>, Supplementary Material S1: Exact and Monte Carlo tests, pooling decisions, and valid n by item; Supplementary Material S2: Results by questionnaire stream; Supplementary Material S3: Component loadings, communalities, and number of readiness indicators answered per respondent; Supplementary Material S4: Robustness check—main questionnaire stream only; Supplementary Material S5: Robustness check—alternative treatment of the undocumented “Good” category; Supplementary Material S6: Robustness check—listwise deletion; Supplementary Material S7: Robustness check—strict seven-of-seven composite inclusion rule; Supplementary Material S8: Questionnaire (Arabic and English versions); Supplementary Material S9: Ordinal models of adoption maturity, Brant tests, and additional robustness results.

Author Contributions: Conceptualization, E.D.; methodology, E.D.; formal analysis, E.D.; investigation, E.D. and A.A.M.; writing—original draft preparation, E.D.; writing—review and editing, E.D. and A.A.M.; project administration, E.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Zayed University (grant number 23228) and United Arab Emirates University (grant number G00005315).

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Research Ethics Committee of Zayed University (protocol code ZU25_074_F).

Informed Consent Statement: Participation was voluntary. Respondents were informed of the study's purpose and the confidentiality of their responses at the start of the questionnaire, and consent was implied by completing the questionnaire.

Data Availability Statement: The anonymized dataset and analysis code are available from the corresponding author on reasonable request. Organization names and open-ended responses are withheld to protect participant anonymity.

Acknowledgments: During the study, the first author used NotebookLM (web-based version, Google, accessed April–August 2026) to assist with organizing and interpreting analysis outputs, and Claude Opus 5.5 (Anthropic, accessed August–September 2026) to assist with revising the manuscript and recomputing the statistical analyses reported in this revision. The authors reviewed and verified all outputs against the data and take full responsibility for the content.

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

References

- Ahmad, N., Haque, S., & Ibahrine, M. (2023). The news ecosystem in the age of AI: Evidence from the UAE. *Journal of Broadcasting & Electronic Media*, 67(3), 323–352. [\[CrossRef\]](#)
- Al-Jenaibi, B., & Bulhoon, A. (2024). Investigating adoption determinants, obstacles, and interventions for AI implementation in Emirati media organizations. *South Eastern European Journal of Public Health*, 25(1), 1901–1909. [\[CrossRef\]](#)
- Almatham, R. N., Darwish, K. M., Al-Rasheed, R., Alshammari, W. T., Alhoshan, M., Almazrui, A., Al Wazrah, A., Alheraki, M., Alam, F., & Nakov, P. (2025). BALSAM: A platform for benchmarking Arabic large language models. In *Proceedings of the third Arabic natural language processing conference* (pp. 258–277). Association for Computational Linguistics.
- Alsheibani, S., Cheung, Y., & Messom, C. (2018). Artificial intelligence adoption: AI-readiness at the firm-level. In *Proceedings of the 22nd Pacific Asia conference on information systems* (Paper 37). Association for Information Systems.
- Bacharach, S. B. (1989). Organizational theories: Some criteria for evaluation. *Academy of Management Review*, 14(4), 496–515. [\[CrossRef\]](#)
- Beckett, C., & Yaseen, M. (2023). *Generating change: A global survey of what news organizations are doing with AI*. Polis, London School of Economics and Political Science.
- Brant, R. (1990). Assessing proportionality in the proportional odds model for ordinal logistic regression. *Biometrics*, 46(4), 1171–1178. [\[CrossRef\]](#)
- Broussard, M., Diakopoulos, N., Guzman, A. L., Abebe, R., Dupagne, M., & Chuan, C.-H. (2019). Artificial intelligence and journalism. *Journalism & Mass Communication Quarterly*, 96(3), 673–695. [\[CrossRef\]](#)
- Busse, C., Kach, A. P., & Wagner, S. M. (2017). Boundary conditions: What they are, how to explore them, why we need them, and when to consider them. *Organizational Research Methods*, 20(4), 574–609. [\[CrossRef\]](#)
- Cools, H., & Diakopoulos, N. (2026). Uses of generative AI in the newsroom: Mapping journalists' perceptions of perils and possibilities. *Journalism Practice*, 20(3), 878–896. [\[CrossRef\]](#)
- Damanpour, F. (1991). Organizational innovation: A meta-analysis of effects of determinants and moderators. *Academy of Management Journal*, 34(3), 555–590. [\[CrossRef\]](#)
- Diakopoulos, N. (2019). *Automating the news: How algorithms are rewriting the media*. Harvard University Press.
- Habash, N. Y. (2010). *Introduction to Arabic natural language processing*. Morgan & Claypool.
- Hassouni, A., & Mellor, N. (2025). AI in the United Arab Emirates' media sector: Balancing efficiency and cultural integrity. *Journalism and Media*, 6(1), 31. [\[CrossRef\]](#)
- Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods*, 7(2), 191–205. [\[CrossRef\]](#)
- Holmström, J. (2022). From AI to digital transformation: The AI readiness framework. *Business Horizons*, 65(3), 329–339. [\[CrossRef\]](#)
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30(2), 179–185. [\[CrossRef\]](#)
- Iacovou, C. L., Benbasat, I., & Dexter, A. S. (1995). Electronic data interchange and small organizations: Adoption and impact of technology. *MIS Quarterly*, 19(4), 465–485. [\[CrossRef\]](#)
- Joshi, P., Santy, S., Budhiraja, A., Bali, K., & Choudhury, M. (2020). The state and fate of linguistic diversity and inclusion in the NLP world. In *Proceedings of the 58th annual meeting of the association for computational linguistics* (pp. 6282–6293). Association for Computational Linguistics.
- Jöhnk, J., Weißert, M., & Wyrski, K. (2021). Ready or not, AI comes: An interview study of organizational AI readiness factors. *Business & Information Systems Engineering*, 63(1), 5–20. [\[CrossRef\]](#)
- Leonardi, P. M. (2011). When flexible routines meet flexible technologies: Affordance, constraint, and the imbrication of human and material agencies. *MIS Quarterly*, 35(1), 147–167. [\[CrossRef\]](#)
- Lewis, S. C., Guzman, A. L., & Schmidt, T. R. (2019). Automation, journalism, and human–machine communication: Rethinking roles and relationships of humans and machines in news. *Digital Journalism*, 7(4), 409–427. [\[CrossRef\]](#)

- Lewis, S. C., Guzmán, A. L., Schmidt, T. R., & Lin, B. (2025). Generative AI and its disruptive challenge to journalism: An institutional analysis. *Communication and Change*, 1(1), 9. [\[CrossRef\]](#)
- Lokuge, S., Sedera, D., Grover, V., & Xu, D. (2019). Organizational readiness for digital innovation: Development and empirical calibration of a construct. *Information & Management*, 56(3), 445–461. [\[CrossRef\]](#)
- Milosavljević, M., & Vobič, I. (2019). Human still in the loop: Editors reconsider the ideals of professional journalism through automation. *Digital Journalism*, 7(8), 1098–1116. [\[CrossRef\]](#)
- Molla, A., & Licker, P. S. (2005). Perceived e-readiness factors in e-commerce adoption: An empirical investigation in a developing country. *International Journal of Electronic Commerce*, 10(1), 83–110. [\[CrossRef\]](#)
- Mousi, B., Durrani, N., Ahmad, F., Hasanain, M., Kabbani, T., Dalvi, F., Chowdhury, S. A., & Alam, F. (2025). AraDiCE: Benchmarks for dialectal and cultural capabilities in LLMs. In *Proceedings of the 31st international conference on computational linguistics* (pp. 4186–4218). Association for Computational Linguistics.
- Møller, L. A., Cools, H., & Skovsgaard, M. (2025). One size fits some: How journalistic roles shape the adoption of generative AI. *Journalism Practice*. Advance online publication. [\[CrossRef\]](#)
- Orlikowski, W. J. (1992). The duality of technology: Rethinking the concept of technology in organizations. *Organization Science*, 3(3), 398–427. [\[CrossRef\]](#)
- Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88(5), 879–903. [\[CrossRef\]](#)
- Rogers, E. M. (2003). *Diffusion of innovations* (5th ed.). Free Press.
- Simon, F. M. (2024). *Artificial intelligence in the news: How AI retools, rationalizes, and reshapes journalism and the public arena*. Tow Center for Digital Journalism, Columbia University.
- Sonni, A. F., Hafied, H., Irwanto, I., & Latuheru, R. (2024). Digital newsroom transformation: A systematic review of the impact of artificial intelligence on journalistic practices, news narratives, and ethical challenges. *Journalism and Media*, 5(4), 1554–1570. [\[CrossRef\]](#)
- Thurman, N., Dörr, K., & Kunert, J. (2017). When reporters get hands-on with robo-writing: Professionals consider automated journalism's capabilities and consequences. *Digital Journalism*, 5(10), 1240–1259. [\[CrossRef\]](#)
- Thurman, N., Thäsler-Kordonouri, S., & Fletcher, R. (2025). *AI adoption by UK journalists and their newsrooms: Surveying applications, approaches, and attitudes* (Report). Reuters Institute for the Study of Journalism. [\[CrossRef\]](#)
- Tornatzky, L. G., & Fleischer, M. (1990). *The processes of technological innovation*. Lexington Books.
- Trist, E. L., & Bamforth, K. W. (1951). Some social and psychological consequences of the longwall method of coal-getting. *Human Relations*, 4(1), 3–38. [\[CrossRef\]](#)
- Whetten, D. A. (1989). What constitutes a theoretical contribution? *Academy of Management Review*, 14(4), 490–495. [\[CrossRef\]](#)

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