journalism and media

MDPI

Article

Curriculum–Skill Gap in the AI Era: Assessing Alignment in Communication-Related Programs

Burak Yaprak [1], Sertaç Ercan [2,*], Bilal Coşan [3] and Mehmet Zahid Ecevit [4]

[1] Department of Business, İstanbul Ticaret University, Istanbul 34445, Türkiye; byaprak@ticaret.edu.tr
[2] Department of Business, Bandirma Onyedi Eylul University, Balikesir 10250, Türkiye
[3] Department of Labor Economics and Industrial Relations, Bandirma Onyedi Eylul University, Balikesir 10250, Türkiye; bcosan@bandirma.edu.tr
[4] Department of Business, Bursa Technical University, Bursa 16310, Türkiye; mehmet.ecevit@btu.edu.tr
* Correspondence: sercan@bandirma.edu.tr

Abstract

Artificial intelligence is rapidly reshaping skill expectations across media, marketing, and journalism, however, university curricula are not evolving at a comparable speed. To quantify the resulting curriculum–skill gap in communication-related programs, two synchronous corpora were assembled for the period July 2024–June 2025: 66 course descriptions from six leading UK universities and 107 graduate-to-mid-level job advertisements in communications, digital media, advertising, and public relations. Alignment around AI, datafication, and platform governance was assessed through a three-stage natural-language-processing workflow: a dual-tier AI-keyword index, comparative TF–IDF salience, and latent Dirichlet allocation topic modeling with bootstrap uncertainty. Curricula devoted 6.0% of their vocabulary to AI plus data/platform terms, whereas job ads allocated only 2.3% ($\chi^2 = 314.4$, $p < 0.001$), indicating a conceptual-critical emphasis on ethics, power, and societal impact in the academy versus an operational focus on SEO, multichannel analytics, and campaign performance in recruitment discourse. Topic modeling corroborated this divergence: universities foregrounded themes labelled "Politics, Power & Governance", while advertisers concentrated on "Campaign Execution & Performance". Environmental and social externalities of AI—central to the Special Issue theme—were foregrounded in curricula but remained virtually absent from job advertisements. The findings are interpreted as an extension of technology-biased-skill-change theory to communication disciplines, and it is suggested that studio-based micro-credentials in automation workflows, dashboard visualization, and sustainable AI practice be embedded without relinquishing critical reflexivity, thereby narrowing the curriculum–skill gap and fostering environmentally, socially, and economically responsible media innovation. With respect to the novelty of this research, it constitutes the first large-scale, data-driven corpus analysis that empirically assessed the AI-related curriculum–skill gap in communication disciplines, thereby extending technology-biased-skill-change theory into this field.

Keywords: technological innovation; artificial intelligence; communication education; curriculum–skill gap; datafication; platform governance; labor-market analytics; natural language processing; topic modeling

check for updates

Academic Editors: Mathias Felipe De-Lima-Santos and Adeola Abdulateef Elega

Received: 7 August 2025
Revised: 15 September 2025
Accepted: 29 September 2025
Published: 6 October 2025

Citation: Yaprak, B., Ercan, S., Coşan, B., & Ecevit, M. Z. (2025). Curriculum–Skill Gap in the AI Era: Assessing Alignment in Communication-Related Programs. *Journalism and Media*, *6*(4), 171. https://doi.org/10.3390/journalmedia6040171

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The proliferation of artificial intelligence (AI) is fundamentally reshaping labor market expectations. Since the public release of ChatGPT in November 2022, vacancies that

stipulate a university degree have fallen by 63% in the United Kingdom and 43% in the United States (Financial Times, 2025). This striking shift is merely the latest phase of a trajectory that began with the Industrial Revolution and has been accelerated by successive waves of digitization and AI driven automation (Ingaldi et al., 2023). While emerging technologies can displace specific roles, they simultaneously create new occupations and skill requirements.

Communication professions—advertising, public relations, marketing communications, and journalism—stand at the epicenter of this transformation. AI, machine learning, and digital workflow automation now redraw the boundaries of professional competence. Employers increasingly expect graduates to pair traditional communication expertise with digital fluency, creative problem solving, and data literacy including skills such as prompt engineering, SEO, A/B testing analytics, and dashboard visualization (Ahadi et al., 2022; WEF, 2025). However, university curricula do not always evolve at the same pace, generating a pronounced 'curriculum gap' that can undermine employability (Vinish & Pinto, 2022). The COVID-19 pandemic also acted as a powerful accelerator of these dynamics, reshaping both labor market structures and higher education delivery. Remote working, digital collaboration platforms, and online teaching rapidly normalized practices that reinforced the demand for hybrid digital–communication skills while exposing structural gaps in university curricula (Crawford et al., 2020; Marinoni et al., 2020).

Recent comparative surveys of journalism programs illustrate how uneven this adaptation can be. In Türkiye, for example, AI-related instruction remains largely theoretical, highly variable from one institution to another, and insufficiently attuned to hands on practice (Babacan et al., 2025). A parallel audit of Spanish journalism degrees likewise found only marginal coverage of AI and big data analytics—just seven full modules explicitly addressed data journalism, while a further 19 merely touched on the topic as part of broader courses (Tejedor et al., 2024). Taken together, these studies suggest that the curriculum gap may be widest precisely where the pace of technological change is fastest.

Although scholarship on digital transformation in media-related subjects is expanding (Westlund et al., 2020), systematic, data driven evaluations of education–workforce alignment remain scarce (Aljohani et al., 2022). The Align My Curriculum (AMC) framework developed by Almaleh et al. (2019) offers a valuable method for automatically comparing syllabi with job postings, but most applications to date focus on computer science rather than communication studies (Ahadi et al., 2022).

This study investigates thematic convergence and divergence between UK communication curricula and labor market demands, focusing specifically on course descriptions from six leading universities. This study analyzed 66 course descriptions from six leading universities alongside 107 job advertisements in digital media, communications, marketing communications, and advertising. Natural language processing (NLP) techniques—AI keyword indexing, TF IDF differential analysis, and latent Dirichlet allocation (LDA) topic modeling—provide a scalable alternative to manual coding frameworks such as AMC. This comparative design gives rise to our central guiding question—later unpacked into three detailed sub questions in the methodology section—namely: To what extent are UK communication related curricula aligned with industry expectations concerning AI and digital technology?

By offering a replicable NLP workflow, the study extends theory of technology biased skill change to communication disciplines and contributes to debates on higher education's responsiveness to labor market needs. Practically, the findings inform actionable recommendations—embedding applied modules and strengthening university–industry partnerships—to bridge the theory–practice divide flagged in recent literature.

This paper proceeds as follows. Section 2 reviews research on AI's environmental, social, and economic impacts and on curriculum alignment frameworks; Section 3 describes the data and methodology; Section 4 presents the results; Section 5 discusses the implications for sustainable, socially responsible, and economically resilient curriculum design, and outlines limitations and future research directions.

This study contributes to the literature by extending technology-biased skill change theory to the field of communication education, where evidence on curriculum–labor market alignment is still limited. Methodologically, it advances a replicable NLP workflow that complements manual frameworks such as AMC. Empirically, it offers one of the first systematic analyses of UK communication curricula vis-à-vis AI-driven labor market demands, thereby informing both scholarship and practice.

2. Theoretical Framework and Literature Review

Industry 4.0 has ushered in a sweeping reconfiguration of work and the competitive employers now expect from graduates. Digital production systems, robotics, and above all, AI, are not only eliminating certain tasks, but also spawning entirely new occupations that demand hybrid skill sets. In this context, lifelong learning, AI and data literacy, analytical and creative thinking, agility, resilience, and human-centered leadership have come to the fore (WEF, 2023). The WEF's (2025) Future of Jobs Report predicts that between 2025 and 2030, roughly seven per cent of current global employment will be reshuffled, with 78 million new jobs created and about 92 million existing roles displaced (WEF, 2025). The fastest growing competencies are projected to be technological—AI, big data, networking, cyber security, and digital literacy—closely followed by socioemotional attributes such as creative thinking, curiosity, and an ethos of continuous learning (Momoh et al., 2025; WEF, 2025).

These aggregate forecasts already register in communication workplaces. Advertising agencies, corporate communication teams, and newsrooms increasingly advertise positions that combine prompt engineering, multichannel analytics, and data-driven storytelling with long standing proficiencies in copy writing, media planning, and campaign strategy. Studies of newsroom practice show that data journalism, once considered a niche specialism, is now a global expectation, but only a minority of journalists feel adequately prepared for it (Heravi & Lorenz, 2020). Classroom research confirms the skills gap: second-year journalism students in Greece reported limited initial understanding of data journalism and only began to appreciate its relevance after a dedicated workshop (Georgiadou & Matsiola, 2023).

Education 4.0 has emerged as the pedagogical counterpart to this technological upheaval, calling for curricula that are flexible, interdisciplinary, and deeply integrated with AI-enabled tools. Scholars advocate digital and cognitive fluency, systems thinking, creativity, collaboration, and active learning methodologies as the guiding principles of contemporary higher education (Moraes et al., 2023; Rabiu et al., 2025; Udvaros et al., 2023). However, implementing these principles in journalism education is complicated by professional norms and values. A Nordic survey of journalism educators, for instance, found strong allegiance to public service ideals such as slow, investigative, and solutions-based journalism, even as the respondents acknowledged the disruptive potential of AI (Jaakkola & Uotila, 2022). This normative stance can slow curricular change, especially where educators lack technical expertise.

The empirical literature consistently shows that universities struggle to translate Education 4.0 principles into practice at a pace commensurate with Industry 4.0. Globally, higher education institutions—technological universities in particular—are criticized for syllabi that lag behind digital economy skill requirements (Karakolis et al., 2022; Tjahjono

et al., 2025). Within communication disciplines, course audits in Turkey and Spain demonstrate a continued emphasis on theoretical treatments of AI and ethics, with comparatively little hands-on training in data scraping, algorithm auditing, or AI-assisted content generation (Babacan et al., 2025; Tejedor et al., 2024). Student-centered experiments suggest that an affirmative, critical data studies approach—one that encourages learners to interrogate platforms and algorithms through creative projects—can enhance data literacy and agency (Sylvia, 2021).

Beyond the classroom, the newsroom remains a crucible for applied innovation. The Global Data Journalism survey highlights not only the technical skills required for data-driven storytelling, but also the ethical and professional values that distinguish journalism from pure data science (Heravi & Lorenz, 2020). These findings dovetail with normative analyses that foreground public interest journalism and democratic accountability (Jaakkola & Uotila, 2022), underscoring the need for curricula that balance technological proficiency with civic responsibility.

Notwithstanding a growing body of evidence, quantitative assessments of the curriculum–industry gap in UK communication programs remain rare. The present study addressed that omission by applying NLP techniques—AI keyword indexing, TF IDF differential analysis, and LDA topic modeling—to 66 course descriptions drawn from 6 leading universities and 107 job advertisements in communications, marketing, digital media, and advertising. By measuring lexical and thematic distances between academic content and labor market signals, the research sought to diagnose the scope and nature of misalignment, inform curriculum designers and policymakers aiming to blend conceptual literacy with applied tool proficiency, and extend technology biased skill shift theory to communication disciplines.

3. Research Methodology

3.1. Research Design

In this study, university curricular discourse and the skills/technology discourse found in job advertisements were examined for alignment around artificial intelligence (AI). AI was treated not as a purely technical add on but as a sociotechnical reconfiguration that reshapes many activities such as newswork and authorship in journalism and media (Diakopoulos, 2019; Lewis et al., 2019; Montal & Reich, 2017). This perspective accords with research on platformization and datafication, which locates AI practices within platform governance, data capture, metrics, and infrastructural control (Gillespie, 2010; Kitchin, 2014; Couldry & Mejias, 2019; Boyd & Crawford, 2012).

To minimize temporal confounds, two synchronous corpora covering the period from July 2024 to June 2025 were compiled. The Curriculum Corpus consisted of 66 course descriptions from six UK universities, whereas the Job Postings Corpus included 107 full-text advertisements retrieved from www.reed.co.uk. The study contributes along three methodological axes. First, AI salience was operationalized through a two tier AI-keyword index: a broad dictionary of 17 lexemes captured AI in conjunction with data/analytics and platform terminology, while a narrow dictionary isolated AI specific terms (e.g., LLM, GPT, transformer, prompt) to test construct validity and robustness to vocabulary choice (Gillespie, 2010; Kitchin, 2014; Couldry & Mejias, 2019). Second, corpus specific vocabularies were characterized by comparative TF–IDF, a standard term weighting method (Salton & Buckley, 1988). Third, curricular thematic structure was recovered with latent Dirichlet allocation (LDA), the number of topics (target k = 6) being selected by c_v coherence, and institution level uncertainty being quantified through B = 2000 document level bootstrap resamples (Blei et al., 2003; Röder et al., 2015; Efron & Tibshirani, 1993).

- RQ 1 (Prevalence difference). Does the AI keyword index—under both broad and narrow specifications—differ significantly between curricula and job advertisements? Differences were evaluated with χ^2 tests; risk ratios (RR) and 95% confidence intervals were computed, and phi (φ) was reported as effect size (Agresti, 2002).
- RQ 2 (Lexical differentiation). Which terms most strongly separate the two discourses? Top TF–IDF terms are reported in the main text, and representative context sentences are provided in the Appendix B (Salton & Buckley, 1988).
- RQ 3 (Thematic differentiation). How are LDA topics (k = 6) distributed across universities, and are observed differences robust under bootstrap uncertainty? Variational inference was used for LDA, k was selected by coherence, and 2.5th–97.5th percentiles of the bootstrap distribution were taken as 95% CIs (Blei et al., 2003; Röder et al., 2015; Efron & Tibshirani, 1993).

This dual dictionary, multimethod strategy allows for the examination of the magnitude (prevalence and effect sizes), content (TF–IDF terms), and structure (LDA topics) of curricular versus labor market discourse, thereby recognizing AI simultaneously as a technical vocabulary and as a data platform regime (Diakopoulos, 2019; Couldry & Mejias, 2019; Kitchin, 2014).

3.2. Data and Corpus Construction

Two synchronized corpora were constructed to enable a like-for-like comparison over the period July 2024–June 2025. The Curriculum Corpus was compiled from publicly available course description pages on the program websites of six UK universities: the London School of Economics, King's College London, the University of Leeds, Goldsmiths, Cardiff University, and the University of Sussex. The Job Postings Corpus was compiled from full-text advertisements retrieved from Reed.co.uk, focusing on categories related to communication, digital media, marketing/advertising, and public relations. Corpus construction followed established corpus linguistic principles including transparent sampling frames, clearly defined inclusion and exclusion criteria, and reproducible harvesting and cleaning procedures (Baker, 2006; McEnery & Hardie, 2012; Sinclair, 1991).

The criteria for content inclusion were determined in advance. For curricula, English-language course descriptions at the undergraduate or postgraduate level were included; duplicate pages, boiler-plate institutional text, and non-course items (e.g., faculty landing pages) were excluded. For job postings, English-language ads with a UK location and graduate-to-mid-level roles were included; internships and duplicate postings were excluded. Deduplication was carried out by URL matching and near-duplicate detection based on title + employer + first 200 characters, an operational rule that accords with labor-market text-mining practice (Hershbein & Kahn, 2018). All records were normalized to UTF-8 and stored with provenance metadata.

The content was based on two primary data sources. The full corpus supports the primary analyses under the broad dictionary, whereas a short-text set—containing condensed curriculum descriptions supplied for this study—serves as a sensitivity test under the narrow dictionary. Corpus sizes are summarized in Table 1. The full corpus underpins the primary analyses, whereas the short-text subset provides a sensitivity check for the narrow dictionary.

Table 1. Corpus summary.

View	Corpus	Documents (n)	Total Tokens
Full (primary)	Curriculum	66	10,938
	Job Ads	107	25,196
Short-text (sensitivity)	Curriculum	60	784
	Job Ads	105	23,043

Note. The short-text subset displays a pronounced token imbalance (784 vs. 23,043). Estimates based on the narrow dictionary were interpreted as sensitivity checks; hence, substantive inferences were grounded in the results from the full corpus.

3.3. Pre-Processing

All texts were normalized to 8, converted to lowercase, and stripped of HTML artefacts, URLs, standalone numerals, and punctuation. An extended English stop list was then applied, after which lemmatization was carried out. Multiword technical expressions (e.g., machine learning, neural network) were captured with pattern rules so that each match was counted as a single unit. Token boundaries were defined to retain intra word hyphens and apostrophes while discarding number heavy tokens. These procedures follow standard corpus linguistic practice that combines transparent sampling with reproducible cleaning, concordancing, and KWIC inspection (Sinclair, 1991; Baker, 2006; McEnery & Hardie, 2012). The model's rule and lookup based lemmatizer provides language specific morphology and stop word flags, thereby ensuring consistent vocabulary reduction before dictionary matching and vectorization (Honnibal & Montani, 2020). Dictionary matches for ai and big were restricted to whole token matches; big was additionally constrained to the bi gram big data in a sensitivity run, which yielded substantively unchanged results (Appendix A).

For downstream modeling, count based representations were preferred for LDA in order to preserve the model's generative assumptions, whereas TF–IDF matrices were produced for comparative term salience and the NMF robustness checks. Count features were created with scikit learn's CountVectorizer, using min_df = 2 to remove extremely rare terms and max_df = 0.90 to exclude corpus wide boiler plate; these hyper parameters constrained the vocabulary to features informative for both topic modeling and comparative weighting (Pedregosa et al., 2011). Scikit learn's Latent Dirichlet Allocation implements variational Bayes; the document–topic prior ($\alpha = 1/k$) and the topic–word prior ($\eta = 0.01$) were tuned as described in Section 3.6.

3.4. AI-Keyword Index: Dictionaries, Rationale, Validation, and Statistical Plan

A two tier dictionary strategy was adopted. A broad dictionary was designed to capture AI together with data/analytics and platform infrastructure terminology, thereby reflecting the embeddedness of AI in datafied, platform governed media systems. In parallel, a narrow dictionary was specified to isolate strictly AI specific terminology and to test construct validity and robustness to vocabulary choice. This design accords with research on platformization and data infrastructures, which emphasizes the coupling of algorithmic techniques with data capture, metrics, and governance (Gillespie, 2010; Kitchin, 2014; Couldry & Mejias, 2019; Boyd & Crawford, 2012).

Broad dictionary (primary measure; 17 lexemes): ai, artificial, algorithm, algorithms, data, digital, machine, learning, analytics, platform, platforms, automation, cloud, big, programmatic, deep, systems.

Narrow dictionary (sensitivity; AI specific): ai; algorithm(s); machine learning; deep learning; neural network(s); transformer(s); llm(s); gpt; nlp; genai; prompt(ing); embedding(s).

For each corpus, the AI-keyword index was defined:

$$AI - Keyword\ Index = \frac{N(AI)}{\mathrm{N(tokens)}}$$

Matches were counted after lemmatization; multiword items were treated as single hits. Whole token matching for ai and big; the big data sensitivity check is documented in Appendix A.

Semantic validation was conducted through a Key Word in Context (KWIC) audit of 50 randomly sampled lines (25 per corpus), double coded by two annotators, and reconciled by adjudication. An overall 94% appropriateness rate was obtained, as summarized in Table 2. χ^2 tests on 2 × 2 contingency tables were used to assess prevalence differences; risk ratios (RR) with 95% confidence intervals and phi (φ) were reported as effect sizes. Because the observed χ^2 statistics were far from the decision boundary, Yates' continuity correction was not applied; sensitivity results can be supplied if required (Agresti, 2002; LibreTexts, 2024).

Table 2. KWIC context validation.

Corpus	Lines (n)	AI-Concordant	Non-Concordant	Validity (%)
Curriculum	25	24	1	96.0
Job Ads	25	23	2	92.0
Total	50	47	3	94.0

Note. Random sampling without replacement; dual coding with adjudication. Full KWIC lines, coding scheme, and inter-coder reconciliation details are provided in Appendix A.

The broad dictionary was used as the primary outcome and was reported for the full corpus. The narrow dictionary was applied to the short-text subset as a sensitivity analysis (Table 3). Directional consistency between the two specifications was taken as evidence that the prevalence findings were robust to dictionary choice.

Table 3. AI-keyword index (narrow dictionary, short-text subset).

Dataset	AI Tokens	Total Tokens	AI Ratio (%)
Curriculum (short)	6	784	0.77
Job Ads	20	23,043	0.09

Note. Difference = 0.68 percentage points (95% CI 0.07–1.29); RR = 8.82 (95% CI 3.55–21.90); $\chi^2(1) = 32.02$, $p < 0.001$; $\varphi = 0.037$. Short-text subset contained 60 curriculum documents (784 tokens) and 105 job ads (23,043 tokens).

3.5. Vectors and Representation

Count based features were generated after pre-processing in order to align with the generative assumptions of LDA. Features were extracted with scikit learn's CountVectorizer, using min_df = 2 to remove hapax like noise and max_df = 0.90 to suppress corpus wide boiler plate. The vectorizer tokenized lowercased, lemmatized forms and produced a sparse CSR matrix suitable for probabilistic modeling (Pedregosa et al., 2011). For comparative salience, a TF–IDF representation was also generated and later used for the TF–IDF tables and the NMF robustness model (Salton & Buckley, 1988).

3.6. Topic Modeling

Topic modeling is a statistical approach used to uncover latent thematic structures within a collection of documents (Nikolenko, 2016). Methods such as topic modeling can help uncover previously unknown linguistic patterns and topics of interest by using algorithms based on word co-occurrences (Incelli, 2025). Latent Dirichlet allocation (LDA)

was fitted with scikit learn's LatentDirichletAllocation, which implements variational Bayes. The batch update method was selected so that each EM step incorporated the entire corpus. Hyper parameters were set to n_components = 6, doc_topic_prior = 1/k, topic_word_prior = 0.01, max_iter = 200, and random_state = 42. Model selection was guided by c_v coherence: k = 4–10 was scanned and k = 6 was retained as the best trade-off between coherence and interpretability. Generalization was evaluated on a 20% stratified held out set, which yielded a perplexity of ≈8.6 (lower values indicate better fit).

Topic labels were assigned by two annotators based on the top 15 terms and representative passages; raw agreement reached 83% and discrepancies were resolved by adjudication (Blei et al., 2003; Röder et al., 2015).

As a robustness check, it was trained on the TF–IDF matrix with the same k = 6. The NMF solution largely overlapped with the LDA themes; divergences are documented in Appendix B with side-by-side top term lists. Because institutions differed in document counts, institution level topic shares were accompanied by 95% bootstrap confidence intervals, obtained by resampling documents within institutions (B = 2000) and taking the 2.5th–97.5th percentiles of the resulting share distributions (Efron & Tibshirani, 1993).

4. Results

4.1. AI-Keyword Coverage

Prevalence differences were evaluated first under the broad dictionary (primary outcome) on the full corpus and subsequently under the narrow dictionary (AI-specific) on the short-text subset as a sensitivity analysis. Differences were tested with χ^2; risk ratios (RR) with 95% confidence intervals and phi (φ) were reported as effect sizes (Agresti, 2002). For the difference in proportions (percentage-point gap), Wald intervals were calculated; alternative interval estimators (e.g., Newcombe, 1998) produced substantively equivalent conclusions.

4.1.1. Full Dictionary (Primary Analysis, Full Corpus)

Curricular and labor market corpora were first compared under the broad dictionary. Token counts and resulting AI-keyword index values are summarized in Table 4.

Curricula exhibited a 3.69 percentage point higher share of AI plus data/platform vocabulary than job advertisements. Figure 1 visualizes these proportions.

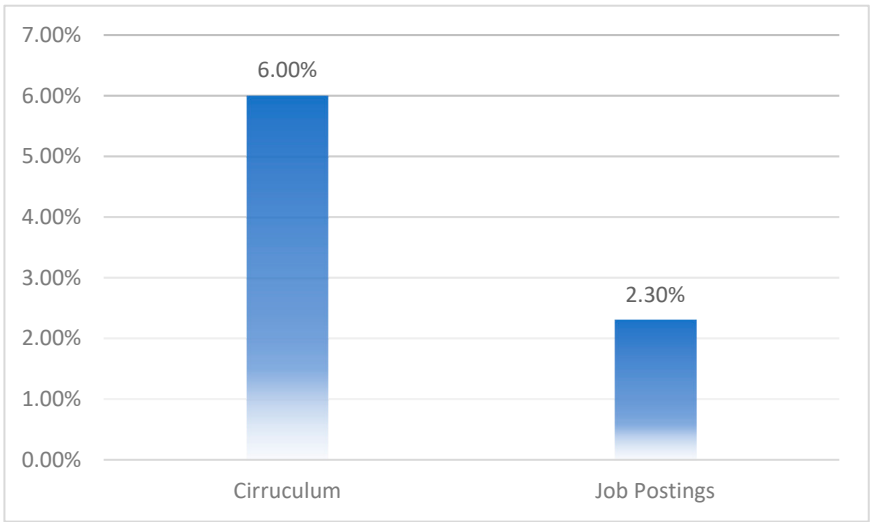


Figure 1. Proportion of AI vs. non-AI tokens in each corpus (broad dictionary, full corpus).

Table 4. AI-keyword index—broad dictionary (full corpus).

Dataset	AI Tokens	Total Tokens	AI Ratio (%)
Curriculum	656	10,938	6.0
Job Postings	581	25,196	2.3

4.1.2. Narrow Dictionary (AI Specific Sensitivity, Short Text Subset)

The narrow dictionary was applied to the short text subset to assess sensitivity to vocabulary choice. Results are summarized in Table 5.

Table 5. AI-keyword index—narrow dictionary (short text subset).

Dataset	AI Tokens	Total Tokens	AI Ratio (%)
Curriculum	6	784	0.77
Job Postings	20	23,043	0.09

A 0.68 percentage point gap was observed. Although absolute rates were low—as expected for a narrow AI lexicon—the direction mirrored the broad dictionary result (curriculum > adverts), supporting robustness to dictionary choice. The pronounced token imbalance (784 vs. 23,043) warrants cautious interpretation of the confidence interval width (Agresti, 2002).

The analysis revealed a statistically significant difference in the frequency of artificial intelligence (AI)-related terminology between university curricula and job postings. The primary analysis, based on a broad AI-related lexicon, showed that the curricula contained 6.0% AI-related terms, whereas job postings included only 2.3% ($\chi^2 = 314.37$, $p < 0.001$; RR = 2.60; $\varphi = 0.093$). This suggests that academic programs make more frequent use of AI-, data-, and platform-related language than is reflected in labor market advertisements. A supplementary sentiment analysis (Figure 2), using a narrower dictionary, yielded lower absolute frequencies (0.77% in curricula vs. 0.09% in job postings), but the directional pattern remained consistent ($\chi^2 = 32.02$, $p < 0.001$; RR = 8.82; $\varphi = 0.037$). These findings indicate that university curricula tend to adopt a more conceptual or critical discourse around AI, whereas job advertisements reflect a narrower, more operational framing of AI-related skills.

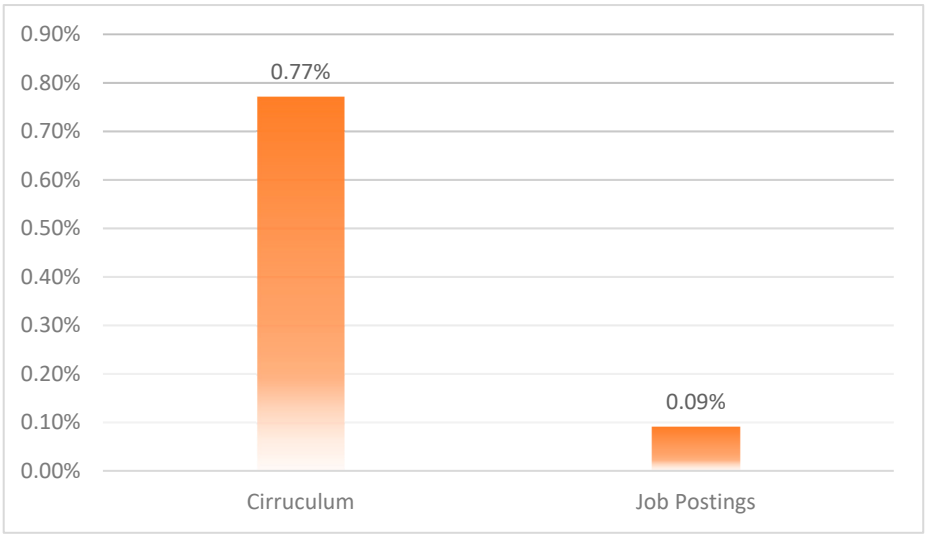


Figure 2. Proportion of AI vs. non-AI tokens in each corpus (narrow dictionary, short-text subset).

4.2. Comparative TF–IDF

Comparative TF–IDF was employed to characterize corpus-specific vocabularies. The weighting scheme down-weights ubiquitous tokens and highlights terms disproportionately characteristic of each corpus, and is widely used for profiling contrasting collections (Salton & Buckley, 1988; Manning et al., 2009). TF–IDF values were computed on lemmatized tokens; results were inspected with KWIC to ensure semantic adequacy. Context sentences are provided in Appendix B (Sinclair, 1991).

The ranking indicates a conceptual/critical vocabulary in curricula (e.g., society, governance, politics, ethics, infrastructure) and an operational/performance vocabulary in job ads (e.g., campaign, SEO, PPC, client, brand, growth). This lexical separation helps explain the prevalence gap observed in Section 4.1: the broad dictionary captures not only AI, but also the surrounding data/metrics/platform discourse that curricula mobilize more systematically, whereas adverts concentrate on channel specific and performance terminology (Gillespie, 2010; Kitchin, 2014; Manning et al., 2009).

A comparative TF–IDF analysis conducted to address the second research question revealed a marked divergence in vocabulary between university curricula and job postings (Table 6). Terms associated with conceptual and critical perspectives—such as media, society, governance, ethics, infrastructure, and politics—were more prevalent in academic course descriptions. In contrast, operational and performance-oriented terms—such as marketing, campaign, client, SEO, PPC, and analytics—dominated job advertisements. This lexical contrast reflects the differing values and priorities of the two discourses: university curricula emphasize a social, structural, and critical engagement with AI and digital technologies, whereas job postings tend to prioritize instrumental skills, channel optimization, and market performance. These findings reinforce the prevalence gap identified in the previous analysis (RQ1), underscoring a thematic misalignment between higher education and industry expectations.

Table 6. Top TF–IDF terms per corpus.

Rank	Curriculum (Conceptual/Critical Register)	Label	Job Ads (Operational/Skills Register)	Label
1	media	Field framing	marketing	Functional role
2	digital	Digital transformation	content	Content production
3	data	Data/analytics lens	brand	Brand management
4	platform	Infrastructure/ecosystem	campaign	Campaign execution
5	society	Societal context	client	Client management
6	governance	Governance/regulation	social	Social media ops
7	politics	Politics/power	performance	Performance metrics
8	power	Power/critique	seo	Search optimization
9	ethics	Accountability	paid	Paid channels
10	method	Research design	ppc	Pay-per-click
11	infrastructure	Technical backbone	analytics	Analytics tools
12	public	Publicness	growth	Growth/scale

4.3. LDA Themes and Institutional Variation (RQ 3)

As preferred by many social science researchers (e.g., Elmholdt et al., 2025), the curricular thematic structure was estimated using latent Dirichlet allocation (LDA) with variational Bayes, as implemented in scikit-learn (Table 7). Topic number was selected by scanning k = 4–10 and retaining k = 6 because of c_v coherence and interpretability. To account for unequal document counts across institutions, institution-level topic shares were accompanied by 95% bootstrap confidence intervals, obtained by resampling documents

within institutions (B = 2000) and taking the 2.5th–97.5th percentiles (Efron & Tibshirani, 1993). Robustness was evaluated with an NMF model trained on TF–IDF features; the resulting themes largely overlapped with the LDA solution (Lee & Seung, 1999).

Table 7. LDA themes (k = 6): Labels, short descriptions, and illustrative top terms.

Rank	Curriculum (Conceptual/Critical Register)	Label	Job Ads (Operational/Skills Register)
T1	Data & Analytics	Data collection, measurement, metrics, analytical literacy	data, analytics, metric, measurement, insight, dashboard, dataset, method
T2	Platforms & Infrastructure	Platform economy, intermediaries, infrastructures, ecosystems	platform, infrastructure, digital, network, system, cloud, service
T3	Politics, Power & Governance	Power, policy, regulation, transparency, accountability	politics, power, governance, regulation, public, policy, accountability
T4	Ethics & Society	Ethics, bias, fairness, social impact	ethics, bias, justice, inclusion, society, harm, responsibility
T5	Methods & Research Design	Research design, evidence, evaluation	method, research, design, empirical, evidence, validity, sample
T6	Audience & Engagement	Audience, participation, community, activism	audience, engagement, community, participation, activism, resistance

Curricular thematic differences across institutions are visualized in Figure 3, which plots the mean topic shares (T1–T6) with their 95% bootstrap confidence intervals (B = 2000). The heat map revealed that Data & Analytics (T1) was especially prominent at Goldsmiths, Audience & Engagement (T6) peaked at King's College, and Politics, Power & Governance (T3) attained higher shares at select institutions. Where intervals overlap, definitive claims are avoided.

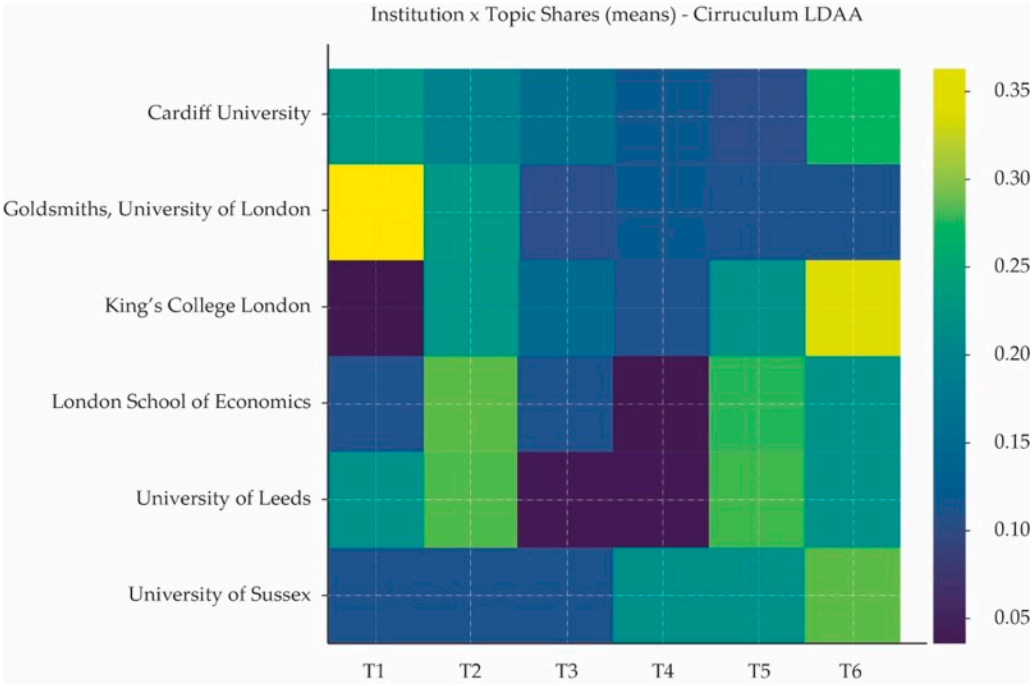


Figure 3. Institution × topic shares with 95% bootstrap confidence intervals.

Cell colors show the mean topic shares; thin bars indicate 95% percentile-bootstrap intervals. Exact values and confidence intervals are reported in Appendix B.

The six themes emerging from the LDA analysis reflect the conceptual diversity and institutional priorities embedded within university curricula related to AI. For instance, Goldsmiths University placed a strong emphasis on Topic 1: Data & Analytics, highlighting terms such as data, analytics, metrics, and dashboard. King's College, in contrast, prioritized Topic 6: Audience & Engagement, as indicated by the frequent use of terms like audience, participation, and community. Other institutions gave prominence to more structural and critical themes, such as Topic 3: Politics, Power & Governance, through terms like power, governance, and policy. This variation illustrates how universities approach AI not merely as a technical subject, but as a phenomenon embedded within broader social, ethical, and political contexts including issues of infrastructure (T2) and ethics and fairness (T4). Collectively, these themes underscore the interdisciplinary nature of AI education and the different ways institutions integrate conceptual, normative, and methodological perspectives into their curricula.

4.4. Synthesis: Areas of Alignment and Gaps

Evidence across Sections 4.1–4.3 indicates that curricula mobilize a conceptual/critical register around AI—linking it to datafication, metrics, platform governance, politics, power, and ethics—whereas job ads emphasize an operational/performance register focused on campaign execution, channel management, SEO/PPC, CRM, and analytics tooling (Gillespie, 2010; Kitchin, 2014; Couldry & Mejias, 2019; Napoli, 2011; Diakopoulos, 2019). These points of alignment and divergence are summarized in Table 8, which maps each curricular theme to its most frequent job-ad counterparts. This pattern is consistent with accounts of automated journalism and human–machine communication that situate AI within reorganized routines and accountability relations, rather than as a narrow technical skill (Lewis et al., 2019; Carlson, 2015; Montal & Reich, 2017; Ananny, 2016).

Table 8. Curriculum themes and job ad skills: alignment matrix.

Curriculum Theme (LDA)	Job Ad Counterparts (High Frequency Terms)	Alignment
Data & Analytics	analytics, measurement, performance, SEO, PPC, dashboard	Partial
Platforms & Infrastructure	platform(s), cloud, CRM, CMS	Partial
Politics, Power & Governance	compliance (rare), policy (rare)	Gap
Ethics & Society	ethics (very rare)	Gap
Methods & Research Design	testing, A/B (limited)	Partial
Audience & Engagement	audience, community, social	Alignment

Pedagogically, the pattern suggests that critical conceptual cores should be preserved, while lab-based, tool-specific modules (e.g., SEO analytics, campaign optimization, CRM workflows, automation) may be introduced as micro-credentials or studio-based courses to bridge operational gaps without compromising critical capacity (Diakopoulos, 2019; Napoli, 2011; Ananny, 2016). Such bridging recognizes AI simultaneously as a technical vocabulary and a data-platform regime, echoing the scholarship on platformization (Gillespie, 2010; Couldry & Mejias, 2019).

When considered collectively, the findings across all research questions indicate that university curricula approach AI through its social, ethical, and managerial dimensions, whereas job postings emphasize technical and operational competencies. This discursive divergence is substantiated by both the TF–IDF and LDA analyses. Although there is some overlap in technical skill areas, notable gaps persist—particularly in themes such as ethics

and governance. To bridge this misalignment, it is recommended that curricula incorporate more practice-oriented modules that address the industry expectations while preserving the critical and conceptual depth of academic content.

4.5. Robustness Checks

Dictionary sensitivity. Directional consistency was observed across specifications: the broad dictionary on the full corpus yielded 6.0% versus 2.3%, whereas the narrow dictionary on the short-text subset yielded 0.77% versus 0.09% (Section 4.1). The larger RR = 8.82 under the narrow dictionary reflects the rarity of AI-specific terms in job ads; interpretation therefore hinges on both absolute and relative magnitudes (Agresti, 2002).

Model and parameter sensitivity. The LDA solution with k = 6 maximized topic coherence in the scanned range; neighboring settings (k = 5/7) regrouped themes without altering the overall semantic backbone. A non-negative matrix factorization (NMF) model trained on TF–IDF features reproduced the major themes, indicating stability across modeling families (Blei et al., 2003; Lee & Seung, 1999; Röder et al., 2015).

Vectorization sensitivity. LDA was fit on count features to respect its generative assumptions; TF–IDF was reserved for comparative tables and the NMF model. Variation in stop-word lists (standard vs. expanded) altered the rank ordering for a small subset of TF–IDF terms but did not change the conceptual/critical versus operational/performance split (Salton & Buckley, 1988; Manning et al., 2009).

Uncertainty estimation. Institution-level topic shares were accompanied by 95% bootstrap confidence intervals; where intervals overlapped substantially, definitive claims were avoided (Efron & Tibshirani, 1993). For χ^2 tests, Yates' continuity correction was not applied given the large χ^2 values, as the correction is recommended primarily for marginal cases near the decision boundary (Agresti, 2002; LibreTexts, 2024).

5. Discussion and Conclusions

The corpus analysis demonstrated a systematic register mismatch between UK communication curricula and labor-market discourse. Course descriptions employ a conceptual–critical vocabulary—datafication, platform governance, ethics—while job advertisements lean on operational keywords—SEO, PPC, multichannel analytics, and client management. Statistically, curricula devoted 6.0% of all tokens to AI + data/platform terms, against 2.3% in recruitment texts ($\chi^2 = 314.4$, $p < 0.001$). Topic modeling corroborated this split: university documents clustered around themes labeled Politics, Power & Governance and Ethics & Society; adverts clustered around Campaign Execution & Performance.

This pattern partially inverts Autor's (2015) technology-biased-skill-change theory. Rather than under-supplying higher-order skills, universities over-supply reflexive competencies and under-supply tool fluency. Employers appear to assume that conceptual awareness can be acquired on the job, whereas specific platform skills must be demonstrated ab initio.

Three mechanisms can explain why critical discourse dominates curricula:

- Disciplinary norms—Journalism and communication educators valorize public-interest functions; ethical and social modules therefore carry curricular weight.
- Accreditation metrics—Program reviews prioritize intellectual rigor over vendor-specific know-how, encouraging theoretical breadth.
- Faculty expertise—Staff publications reside mainly in critical media studies; hold current certifications in tools such as GA4, HubSpot, or programmatic ad platforms.

The findings suggest that alignment can be improved without sacrificing critical depth:

- Micro-credentials—Short, stackable badges in prompt engineering, dashboard visualization and A/B testing can inject operational skills rapidly.

- AI studios—Capstone "clinics" pairing students with industry mentors can merge reflexive inquiry with tool practice.
- Living syllabi—Git-managed curricula updated each semester by joint faculty-practitioner boards can keep pace with fast-moving platforms.
- Impact-focused assessment—Grading rubrics that reward both ethical reasoning and measurable optimization outcomes encourage balanced skill sets.

At the same time, the scope of this study was deliberately narrow: it was confined to six UK universities and a single job board. This limitation means that the conclusions should be read as indicative rather than definitive. Although dictionary sensitivity checks were run, AI jargon evolves quickly; lexicons will require periodic revision. A longitudinal extension could reveal whether curricula and adverts converge over time, while cross-country corpora would test cultural moderators of the gap. Finally, linking lexical alignment scores to graduate employment data would provide behavioral validation of the textual findings.

A scalable corpus-linguistic workflow was employed to diagnose the curriculum–skill gap in communication-related programs during the first post-ChatGPT academic year. Results indicate that universities emphasize conceptual and ethical dimensions of AI, whereas employers foreground platform-specific, performance-oriented skills. This divergence resonates with Autor's technology-biased-skill-change theory, which posits that technological advances systematically shift the relative demand for cognitive versus routine skills. Bridging this register mismatch calls for studio-based, micro-credentialed interventions that embed operational tool practice within a critical pedagogical framework. Doing so would produce graduates who both know why and know how, aligning academic outputs with an AI-transformed labor market while preserving the reflective capacities that differentiate communication professions from purely technical fields.

Author Contributions: Conceptualization, B.Y., S.E., B.C., and M.Z.E.; Methodology, B.Y. and S.E.; Data curation, B.Y.; Writing—original draft preparation, M.Z.E. and B.C.; Writing—review and editing, B.Y., S.E., B.C., and M.Z.E.; Visualization, S.E.; Supervision, B.Y. and B.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

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

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial intelligence
SEO	Search engine optimization
AMC	Align My Curriculum
NLP	Natural language processing
LDA	Latent Dirichlet allocation
CBI	Confederation of British Industry
WEF	World Economic Forum
NMF	Non-negative matrix factorization
UK	United Kingdom
UTF	Unicode Transformation Format

KWIC	Key Word in Context
RR	Risk ratios
PPC	Pay-per-click

Appendix A. Pre-Processing Resources and Validation Details

Appendix A.1. Extended Stop-Word List

The standard spaCy English stop-word list (v3.7.2) was used and augmented with the following domain-neutral tokens: course, module, program, credit, assessment, applicants, apply.

Appendix A.2. Multi-Word Expression (MWE) Patterns

The following MWEs were matched as single tokens via regex or spaCy Matcher:

- Machine learning;
- Neural network;
- Deep learning;
- Big data;
- Natural language processing;
- Large language model;
- Content management system.

Appendix A.3. Token-Level Constraints

Dictionary matches for AI-related tokens followed these rules:

- ai $\rightarrow$ whole-token match (regex\bai\b) to avoid substrings (e.g., brain, chair);
- big $\rightarrow$ whole-token match; in sensitivity analysis restricted to bi-gram big data;
- Results under big data restriction were substantively unchanged (see Section 4.).

Appendix A.4. KWIC Validation Protocol

A simple-random sample of 50 dictionary hits (25 per corpus) was extracted using spaCy's Doc.bin API. Two independent coders labelled each hit as semantically appropriate or not. Disagreements were discussed until consensus was achieved (94% final agreement). Cohen's κ was not computed due to class imbalance.

Appendix A.5. Pre-Processing Pipeline (Python Snippet)

```
import spacy, re
from spacy.matcher import PhraseMatcher

nlp = spacy.load("en_core_web_sm", disable = ["ner", "parser"])
extra_stops = {...}  # see Section A
for t in extra_stops: nlp.vocab[t].is_stop = True

mwes = ["machine learning", "neural network", "big data", "..."]
matcher = PhraseMatcher(nlp.vocab, attr = "LOWER")
patterns = [nlp.make_doc(t) for t in mwes]
matcher.add("MWE", patterns)

def preprocess(text):
    doc = nlp(text.lower())
    matches = matcher(doc)  # merge MWEs
    with doc.retokenize() as retok:
```

```
        for _, start, end in matches:
            retok.merge(doc[start:end])
    tokens = [t.lemma_ for t in doc if not (t.is_stop or t.is_punct or t.like_num)]
    return tokens
```

Appendix B. TF–IDF Details and Context Sentences

Appendix B.1. Top TF–IDF Terms (Excerpt)

The full TF–IDF rankings (n = 30 terms per corpus) are available upon request. Table A1 lists the top 12 terms per corpus for illustration. Numerical scores were rounded.

Table A1. (**a**) Curriculum Corpus—top TF–IDF terms. (**b**) Job-Ads Corpus—top TF–IDF terms.

(**a**)		
Rank	**Term**	**TF–IDF Score**
1	media	—
2	digital	—
3	data	—
4	platform	—
5	society	—
6	governance	—
7	politics	—
8	power	—
9	ethics	—
10	method	—
11	infrastructure	—
12	public	—
(**b**)		
Rank	**Term**	**TF–IDF Score**
1	marketing	—
2	content	—
3	brand	—
4	campaign	—
5	client	—
6	social	—
7	performance	—
8	seo	—
9	paid	—
10	ppc	—
11	analytics	—
12	growth	—

Appendix B.2. KWIC Context Sentences (Top Terms)

Table A2. KWIC context sentences (top terms).

Term	**Corpus**	**KWIC (±5 Tokens)**
society	Curriculum	. . . shaping society through algorithmic news distribution . . .
seo	Job Ads	. . . experience in SEO and content optimization required . . .

Appendix B.3. Sampling and Validation Procedure

KWIC lines were sampled using a uniform random draw without replacement. Two annotators independently judged semantic appropriateness; raw agreement was 94%. Disagreements were resolved by discussion.

References

Agresti, A. (2002). *Categorical data analysis* (2nd ed.). Wiley.

Ahadi, A., Kitto, K., Rizoiu, M. A., & Musial-Gabrys, K. (2022, July 24–27). *Skills taught vs skills sought: Using skills analytics to identify the gaps between curriculum and job markets*. International Conference on Educational Data Mining, Durham University, Durham, UK.

Aljohani, N. R., Aslam, A., Khadidos, A. O., & Hassan, S. U. (2022). Bridging the skill gap between the acquired university curriculum and the requirements of the job market: A data-driven analysis of scientific literature. *Journal of Innovation & Knowledge*, *7*(3), 100190. [CrossRef]

Almaleh, A., Aslam, M. A., Saeedi, K., & Aljohani, N. R. (2019). Align My Curriculum: A framework to bridge the gap between acquired university curriculum and required market skills. *Sustainability*, *11*(9), 2607. [CrossRef]

Ananny, M. (2016). Toward an ethics of algorithms: Convening, observation, probability, and timeliness. *Science, Technology, & Human Values*, *41*(1), 93–117.

Autor, D. H. (2015). Why are there still so many jobs? The history and future of workplace automation. *Journal of Economic Perspectives*, *29*(3), 3–30. [CrossRef]

Babacan, H., Arık, E., Bilişli, Y., Akgün, H., & Özkara, Y. (2025). Artificial intelligence and journalism education in higher education: Digital transformation in undergraduate and graduate curricula in Türkiye. *Journalism and Media*, *6*(2), 52. [CrossRef]

Baker, P. (2006). *Using corpora in discourse analysis*. Continuum. A&C Black.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, *3*, 993–1022.

Boyd, d., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, *15*(5), 662–679. [CrossRef]

Carlson, M. (2015). The robotic reporter: Automated journalism and the redefinition of labor, compositional forms, and journalistic authority. *Digital Journalism*, *3*(3), 416–431. [CrossRef]

Couldry, N., & Mejias, U. A. (2019). *The costs of connection: How data is colonizing human life and appropriating it for capitalism*. Stanford University Press.

Crawford, J., Butler-Henderson, K., Rudolph, J., Malkawi, B., Glowatz, M., Burton, R., Magni, P., & Lam, S. (2020). COVID-19: 20 countries' higher education intra-period digital pedagogy responses. *Journal of Applied Learning & Teaching*, *3*(1), 1–20. [CrossRef]

Diakopoulos, N. (2019). *Automating the news: How algorithms are rewriting the media*. Harvard University Press.

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. Chapman & Hall/CRC.

Elmholdt, K. T., Nielsen, J. A., Florczak, C. K., Jurowetzki, R., & Hain, D. (2025). The hopes and fears of artificial intelligence: A comparative computational discourse analysis. *AI & Society*, *40*, 4765–4782. [CrossRef]

Financial Times. (2025). *Is AI killing graduate jobs?* Available online: https://www.ft.com/content/99b6acb7-a079-4f57-a7bd-8317c1fbb728 (accessed on 28 July 2025).

Georgiadou, E., & Matsiola, M. (2023). Understanding and enhancing journalism students' perception of data journalism. *Journalism and Media*, *4*(4), 1232–1247. [CrossRef]

Gillespie, T. (2010). The politics of "platforms". *New Media & Society*, *12*(3), 347–364. [CrossRef]

Heravi, B. R., & Lorenz, M. (2020). Data journalism practices globally: Skills, education, opportunities, and values. *Journalism and Media*, *1*(1), 26–40. [CrossRef]

Hershbein, B., & Kahn, L. B. (2018). Do recessions accelerate routine-biased technological change? *American Economic Review*, *108*(7), 1737–1772. [CrossRef]

Honnibal, M., & Montani, I. (2020). spaCy 2: Natural language understanding with Bloom embeddings and convolutional neural networks. In *Proceedings of the 7th workshop on NLP open source software* (pp. 10–18). Association for Computational Linguistics. [CrossRef]

Incelli, E. (2025). Exploring the future of corpus linguistics: Innovations in AI and social impact. *International Journal of Mass Communication*, *3*, 1–10. [CrossRef]

Ingaldi, M., Ulewicz, R., & Klimecka-Tatar, D. (2023). Creation of the university curriculum in the field of Industry 4.0 with the use of modern teaching instruments-Polish case study. *Procedia Computer Science*, *217*, 660–669. [CrossRef]

Jaakkola, M., & Uotila, P. (2022). Exploring the normative foundation of journalism education: Nordic journalism educators' conceptions of future journalism and professional qualifications. *Journalism and Media*, *3*(3), 436–452. [CrossRef]

Karakolis, E., Kapsalis, P., Skalidakis, S., Kontzinos, C., Kokkinakos, P., Markaki, O., & Askounis, D. (2022). Bridging the gap between technological education and job market requirements through data analytics and decision support services. *Applied Sciences*, *12*(14), 7139. [CrossRef]

Kitchin, R. (2014). *The data revolution: Big data, open data, data infrastructures & their consequences*. SAGE.

Lee, D. D., & Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, *401*, 788–791. [CrossRef]

Lewis, S. C., Guzman, A. L., & Schmidt, T. R. (2019). Automation, journalism, and human–machine communication: Responding to the co-evolution of social and technological change. *Digital Journalism*, *7*(4), 409–427. [CrossRef]

LibreTexts. (2024). *Yates' continuity correction*. Available online: https://stats.libretexts.org/ (accessed on 30 July 2025).

Manning, C. D., Raghavan, P., & Schütze, H. (2009). *Introduction to information retrieval*. Cambridge University Press.

Marinoni, G., Van't Land, H., & Jensen, T. (2020). *The impact of COVID-19 on higher education around the world*. International Association of Universities (IAU). Available online: https://www.iau-aiu.net/IMG/pdf/iau_covid19_and_he_survey_report_final_may_2020.pdf (accessed on 12 July 2025).

McEnery, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.

Momoh, A. M., Olajide, F. O., Ogundipe, R. O., & Adesina, A. D. (2025). Integration of artificial intelligence in industrial education: A review of current trends and future directions. *Journal of Computer, Software, and Program*, *2*(2), 1–9. [CrossRef]

Montal, T., & Reich, Z. (2017). I, robot. You, journalist. Who is the author? *Digital Journalism*, *5*(7), 829–849. [CrossRef]

Moraes, E. B., Kipper, L. M., Hackenhaar Kellermann, A. C., Austria, L., Leivas, P., Moraes, J. A. R., & Witczak, M. (2023). Integration of Industry 4.0 technologies with Education 4.0: Advantages for improvements in learning. *Interactive Technology and Smart Education*, *20*(2), 271–287. [CrossRef]

Napoli, P. M. (2011). *Audience evolution: New technologies and the transformation of media audiences*. Columbia University Press.

Newcombe, R. G. (1998). Interval estimation for the difference between independent proportions: Comparison of eleven methods. *Statistics in Medicine*, *17*(8), 873–890. [CrossRef]

Nikolenko, S. I. (2016, July 17–21). *Topic quality metrics based on distributed word representations*. 39th International ACM SIGIR conference on Research and Development in Information Retrieval (pp. 1029–1032), Pisa, Italy. [CrossRef]

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Rabiu, A., Bawa, K., & Saminu, S. (2025). Artificial intelligence adoption for skills development in Nigeria: A systematic review and roadmap for TVET transformation. *International Journal of Research and Innovation in Social Science*, *9*(8), 645–675. [CrossRef]

Röder, M., Both, A., & Hinneburg, A. (2015, February 2–6). *Exploring the space of topic coherence measures*. Eighth ACM International Conference on Web Search and Data Mining (pp. 399–408), Shanghai, China. [CrossRef]

Salton, G., & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *Information Processing & Management*, *24*(5), 513–523. [CrossRef]

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.

Sylvia, J. J., IV. (2021). An affirmative approach to teaching critical data studies. *Journalism and Media*, *2*(4), 641–656. [CrossRef]

Tejedor, S., Cervi, L., Romero-Rodríguez, L. M., & Vick, S. (2024). Integrating artificial intelligence and big data in Spanish journalism education: A curricular analysis. *Journalism and Media*, *5*(4), 1607–1623. [CrossRef]

Tjahjono, B., Hermawan, D., Millah, S., & Evans, R. (2025). Bridging the skills gap curriculum transformation for automation industries and the role of digital technopreneurship. *Aptisi Transactions on Technopreneurship*, *7*(2), 650–662. [CrossRef]

Udvaros, J., Gubán, M., Gubán, Á., & Sándor, Á. (2023). Industry 4.0 from the perspective of Education 4.0. *International Journal of Advanced Natural Sciences and Engineering Researches*, *7*(4), 230–234. [CrossRef]

Vinish, P., & Pinto, P. (2022). Framework for identification of curriculum gaps: A systematic approach. *Journal of Engineering Education Transformations*, *35*, 61–68. [CrossRef]

Westlund, O., Krumsvik, A. H., & Lewis, S. C. (2020). Competition, change, and coordination and collaboration: Tracing news executives' perceptions about participation in media innovation. *Journalism Studies*, *22*(1), 1–21. [CrossRef]

World Economic Forum. (2023). *Future of jobs report 2023*. Available online: https://www3.weforum.org/docs/WEF_Future_of_Jobs_2023.pdf (accessed on 9 July 2025).

World Economic Forum. (2025). *Future of jobs report 2025*. Available online: https://reports.weforum.org/docs/WEF_Future_of_Jobs_Report_2025.pdf (accessed on 9 July 2025).

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.