

I

CHAPTER ONE

HUMAN RESOURCES AND ARTIFICIAL INTELLIGENCE: A TEXTUAL ANALYSIS OF EMERGING RESEARCH DIRECTIONS

Prof. Dr. Milos Petkovic

Human Resources and Artificial Intelligence: A Textual Analysis of Emerging Research Directions

*Milos Petkovic*¹

¹ SBS Swiss Business School, Zurich, Switzerland
Corresponding Author: Milos Petkovic, milos.petkovic@sbs.edu

Chapter Information

- **Date of Receipt:** [01/03/2026]
- **Date of Acceptance:** [01/04/2026]
- **JEL Classification Codes :** J24 ; O33 ; C38 ;

Abstract

This study maps the dominant themes in recent research at the intersection of human resources, AI, and machine learning, based on 28 journal articles from applied science journals published over the past year. Using IRaMuTeQ software, it applied Reinert hierarchical classification and Correspondence Factor Analysis, producing visualizations and matrices that identified four key thematic clusters: AI Adoption in Education, Healthcare & Organizations (28.8%), ML Model Performance & Evaluation (28.8%), Automated Image Classification & Computer Vision (18.9%), and Geospatial AI & System Monitoring (23.4%). The factorial plan explained 53.5% of the variance, indicating strong independence among the clusters. Findings show AI/ML research expanding into healthcare, education, environmental monitoring, and infrastructure, urging human resource professionals and organizational leaders to consider cross-disciplinary AI applications. Limitations include the focus on 28 articles from a single year and applied science journals, which may omit broader perspectives. The analysis reflects lexical patterns, not deeper semantics. This overview offers a data-driven map of current AI/ML and HR research intersections, contributing methodologically to bibliometric and textual studies in management and sciences.

Keywords: *artificial intelligence; machine learning; human resources; textual analysis.*

INTRODUCTION

The rapid advancement of artificial intelligence and machine learning technologies has initiated a profound transformation across virtually every sector of the global economy, reshaping organizational structures, redefining professional roles, and fundamentally altering the nature of human work. What was once confined to the realm of computer science and engineering laboratories has become a pervasive organizational reality, penetrating strategic decision-making, operational processes, and human resource management practices at an unprecedented pace. This technological revolution presents both extraordinary opportunities and significant challenges for organizations, policymakers, researchers, and the workforce alike, making the systematic study of AI and machine learning developments not merely an academic exercise but an urgent practical necessity.

Within this broader context, human resource management occupies a particularly sensitive and strategically significant position. As the organizational function most directly responsible for attracting, developing, managing, and retaining human talent, HRM stands at the epicenter of the disruption generated by AI and machine learning adoption. The questions being asked today — How will AI reshape job roles and required competencies? How can organizations harness predictive analytics to improve workforce decisions? What ethical frameworks should govern algorithmic HR tools? — are among the most consequential in contemporary management discourse. Yet despite the growing volume of scholarly output addressing these questions, a coherent and empirically grounded map of where the research field currently stands remains elusive, particularly when viewed through the lens of applied science literature.

The last twelve months have witnessed an exceptional acceleration in both the development and deployment of AI technologies, including large language models, generative AI systems, computer vision applications, and geospatial monitoring platforms. This acceleration has been mirrored in the academic literature, with applied science journals publishing an increasing volume of research exploring AI and machine learning applications across healthcare, education, environmental science, agriculture, and organizational management. However, the sheer diversity and volume of this output makes it difficult for scholars and practitioners to identify the dominant thematic directions, emerging research clusters, and knowledge gaps that characterize the field at any given moment. This challenge calls for structured, methodologically rigorous approaches capable of synthesizing large volumes of textual information into interpretable thematic insights.

It is precisely this gap that the present study seeks to address. By applying IRaMuTeQ-based qualitative textual analysis to a corpus of 28 peer-reviewed journal articles published in applied science journals over the past twelve months, this chapter aims to identify and map the dominant thematic clusters emerging at the intersection of human resources, artificial intelligence, and machine learning. Rather than offering a traditional narrative literature review, this study adopts a corpus-based analytical approach that leverages the statistical power of Reinert hierarchical classification and Correspondence Factor Analysis to surface latent thematic structures that might not be visible through conventional review methods. In doing so, it contributes both substantively — by identifying current research directions — and methodologically — by demonstrating the applicability of textual analysis tools to the management and applied science literature.

The analysis produced four statistically distinct thematic clusters, collectively accounting for all 111 text segments analyzed. The first cluster, representing 28.8% of the corpus, centers on AI adoption in educational, healthcare, and organizational contexts, reflecting the growing scholarly attention to how individuals and institutions experience and integrate AI

tools into their daily practices. The second cluster, equally weighted at 28.8%, addresses machine learning model performance and predictive evaluation, capturing the dominant strand of technical research concerned with benchmarking and optimizing ML algorithms across applied domains. The third cluster, comprising 18.9% of the corpus, focuses on automated image classification and computer vision, revealing a specialized but significant body of research applying deep learning to biological, agricultural, and medical imaging challenges. The fourth cluster, representing 23.4% of the corpus, encompasses geospatial AI and real-world system monitoring, highlighting the growing application of AI to large-scale environmental, infrastructural, and epidemiological surveillance.

Together, these four clusters trace a compelling trajectory in the contemporary literature — from the micro-level concerns of individual AI adoption and organizational impact, through the methodological preoccupations of model performance and technical automation, to the macro-level applications of AI in monitoring and managing complex global systems. This trajectory reflects not only the broadening scope of AI research but also its deepening integration into the practical fabric of human activity across an expanding range of scientific and organizational domains.

The remainder of this chapter is structured as follows. The Literature Review section synthesizes existing scholarship across the three core thematic domains underpinning this study: AI and human resource management, machine learning applications in organizational and applied science research, and textual and bibliometric analysis as a research methodology. The Research Methodology section describes the corpus construction process, the IRaMuTeQ analytical framework, and the validity and reliability considerations governing the study. The Findings section presents and interprets the results of the textual analysis, including the dendrogram, factorial plan, semantic cloud, word cloud, and Labbé distance matrix. The Discussion section situates the findings within the broader scholarly conversation, drawing out theoretical and practical implications. The Conclusion section summarizes the key contributions of the study, acknowledges its limitations, and proposes directions for future research.

LITERATURE REVIEW

The intersection of human resources management, artificial intelligence, and machine learning represents one of the most rapidly evolving areas of organizational and applied science research. Over the past decade, scholars have progressively shifted their attention from theoretical explorations of AI potential to empirical investigations of its practical implications for workforce management, organizational performance, and technological innovation. This literature review synthesizes the existing body of knowledge across three interconnected thematic domains: the transformation of human resource management through artificial intelligence; machine learning applications in organizational and applied science contexts; and the methodological evolution of textual and bibliometric analysis in management research.

Artificial Intelligence and Human Resource Management

The integration of artificial intelligence into human resource management (HRM) has fundamentally altered how organizations recruit, develop, evaluate, and retain talent. Foundational scholarship in this area established that AI-driven technologies are not merely administrative tools but strategic assets capable of reshaping the entire employment lifecycle (Brynjolfsson & McAfee, 2014). The displacement and augmentation debate — whether AI replaces

or enhances human workers — has been central to this discourse, with evidence suggesting that the relationship is highly context-dependent and mediated by organizational readiness, workforce skill levels, and the nature of tasks performed (Acemoglu & Restrepo, 2018). In the domain of talent acquisition, AI-powered recruitment systems have demonstrated significant capacity to reduce hiring time, improve candidate screening accuracy, and minimize unconscious human bias, though scholars have simultaneously raised concerns about algorithmic discrimination and the erosion of human judgment in hiring decisions (Raghavan et al., 2020). Predictive analytics tools, powered by machine learning, have been increasingly deployed in performance management and employee retention, enabling HR professionals to anticipate turnover risk and intervene proactively (Sajjadi et al., 2019). These developments align with the broader theoretical argument that AI augments rather than replaces strategic HR decision-making, particularly in complex, judgment-intensive tasks (Daugherty & Wilson, 2018). The adoption of conversational AI and large language models in organizational settings has introduced new dimensions to the HR-AI relationship. Chatbots and virtual assistants are increasingly embedded in onboarding processes, employee self-service platforms, and learning management systems, reshaping the employee experience and the administrative role of HR departments (Vrontis et al., 2022). Scholars have also highlighted the critical importance of digital literacy and AI competency development as emerging strategic priorities for HR functions, arguing that workforce upskilling is essential to realizing the productivity potential of AI adoption (Huang & Rust, 2021). Furthermore, ethical governance frameworks for AI in HRM have gained prominence, with researchers emphasizing the need for transparency, accountability, and human oversight in algorithmic decision-making processes that affect employees (Tambe et al., 2019).

The intersection of human resources management, artificial intelligence, and machine learning represents one of the most rapidly evolving areas of organizational and applied science research. Over the past decade, scholars have progressively shifted their attention from theoretical explorations of AI potential to empirical investigations of its practical implications for workforce management, organizational performance, and technological innovation. This literature review synthesizes the existing body of knowledge across three interconnected thematic domains: the transformation of human resource management through artificial intelligence; machine learning applications in organizational and applied science contexts; and the methodological evolution of textual and bibliometric analysis in management research.

Artificial Intelligence and Human Resource Management

The integration of artificial intelligence into human resource management (HRM) has fundamentally altered how organizations recruit, develop, evaluate, and retain talent. Foundational scholarship in this area established that AI-driven technologies are not merely administrative tools but strategic assets capable of reshaping the entire employment lifecycle (Brynjolfsson & McAfee, 2014). The displacement and augmentation debate — whether AI replaces or enhances human workers — has been central to this discourse, with evidence suggesting that the relationship is highly context-dependent and mediated by organizational readiness, workforce skill levels, and the nature of tasks performed (Acemoglu & Restrepo, 2018).

In the domain of talent acquisition, AI-powered recruitment systems have demonstrated significant capacity to reduce hiring time, improve candidate screening accuracy, and minimize unconscious human bias, though scholars have simultaneously raised concerns about algorithmic discrimination and the erosion of human judgment in hiring decisions (Raghavan et al., 2020). Predictive analytics tools, powered by machine learning, have been increasingly

deployed in performance management and employee retention, enabling HR professionals to anticipate turnover risk and intervene proactively (Sajjadi et al., 2019). These developments align with the broader theoretical argument that AI augments rather than replaces strategic HR decision-making, particularly in complex, judgment-intensive tasks (Daugherty & Wilson, 2018).

The adoption of conversational AI and large language models in organizational settings has introduced new dimensions to the HR-AI relationship. Chatbots and virtual assistants are increasingly embedded in onboarding processes, employee self-service platforms, and learning management systems, reshaping the employee experience and the administrative role of HR departments (Vrontis et al., 2022). Scholars have also highlighted the critical importance of digital literacy and AI competency development as emerging strategic priorities for HR functions, arguing that workforce upskilling is essential to realizing the productivity potential of AI adoption (Huang & Rust, 2021). Furthermore, ethical governance frameworks for AI in HRM have gained prominence, with researchers emphasizing the need for transparency, accountability, and human oversight in algorithmic decision-making processes that affect employees (Tambe et al., 2019).

Machine Learning Applications in Organizational and Applied Science Research

Machine learning has emerged as a transformative methodological and operational paradigm across both organizational research and applied scientific disciplines. At its core, machine learning encompasses a family of algorithms — including supervised learning, unsupervised learning, deep learning, and reinforcement learning — that enable systems to learn patterns from data without being explicitly programmed, progressively improving their predictive accuracy through experience (Hastie et al., 2009; Mitchell, 1997). The organizational implications of this technological shift are profound, as machine learning enables the extraction of actionable insights from data volumes and complexities that exceed traditional analytical capacities (Shmueli et al., 2016).

In organizational research, machine learning methods have been applied to a wide range of problems including demand forecasting, supply chain optimization, customer behavior prediction, and financial risk modeling (Athey & Imbens, 2019). The growing availability of large organizational datasets — generated by enterprise resource planning systems, customer relationship management platforms, and digital communication tools — has created fertile ground for ML-driven organizational analytics. Scholars have argued that machine learning represents a paradigmatic shift in organizational science methodology, enabling inductive, data-driven theory generation alongside traditional deductive approaches (Agrawal et al., 2018).

In applied science contexts, machine learning has demonstrated transformative impact across healthcare, environmental science, agriculture, and infrastructure management. In healthcare, deep learning models have achieved diagnostic accuracy comparable to specialist clinicians in areas such as medical imaging, pathology, and genomics, fundamentally challenging traditional models of clinical decision-making (Topol, 2019). In environmental and geospatial sciences, ML algorithms applied to satellite imagery and sensor data have enabled real-time monitoring of climate change indicators, natural disaster risk, and urban infrastructure degradation at scales previously unattainable (Reichstein et al., 2019). In agriculture, computer vision systems powered by convolutional neural networks have automated crop disease detection, yield estimation, and precision farming interventions, contributing directly to global food security objectives (Kamilaris & Prenafeta-Boldú, 2018).

The convergence of machine learning with human resource concerns is particularly ev-

ident in workforce analytics, where predictive models are used to optimize staffing decisions, identify high-potential employees, and model the organizational impact of workforce restructuring scenarios (Tursunbayeva et al., 2018). This convergence underscores the argument that machine learning is not a domain-specific tool but a general-purpose technology whose organizational and societal implications are still being fully understood (Brynjolfsson et al., 2017).

Textual and Bibliometric Analysis as a Research Methodology in Management

The application of textual and bibliometric analysis methods to management and organizational research has grown substantially over the past two decades, driven by the increasing availability of large digital text corpora and the development of accessible computational tools for their analysis. Bibliometric methods — including citation analysis, co-authorship mapping, and keyword co-occurrence analysis — have become standard instruments for mapping the intellectual structure of research fields, identifying emerging trends, and tracing the evolution of scholarly discourse (Donthu et al., 2021). These methods offer a systematic and replicable alternative to traditional narrative literature reviews, particularly in rapidly evolving fields where the volume of publications makes comprehensive manual synthesis impractical.

Qualitative textual analysis, as distinct from bibliometrics, focuses on the lexical and semantic content of texts rather than their citation relationships. Software tools such as IRaMuTeQ, NVivo, ATLAS.ti, and MAXQDA have enabled researchers to apply statistical rigor to the analysis of interview transcripts, policy documents, corporate reports, and scientific literature, bridging the methodological gap between qualitative interpretation and quantitative analysis (Reinert, 1990). The Reinert method in particular, which underlies IRaMuTeQ's hierarchical classification algorithm, has been validated across numerous disciplines as a reliable technique for identifying latent thematic structures in textual corpora (Camargo & Justo, 2013).

In management research, corpus-based textual analysis has been applied to studies of organizational discourse, strategic communication, corporate social responsibility reporting, and knowledge management (Gavard-Perret & Moscarola, 1998). Scholars have argued that these methods offer unique epistemological advantages in organizational research, as they allow the researcher to surface implicit assumptions, dominant narratives, and ideological structures embedded in organizational texts that would not be visible through conventional content analysis (Lebart & Salem, 1994). The application of factorial correspondence analysis to textual data — as performed in the present study — provides a spatial representation of semantic relationships that complements the hierarchical structure revealed by cluster analysis, offering a richer and more multidimensional understanding of the corpus (Benzécri, 1992).

More recently, the integration of machine learning techniques — including natural language processing, topic modeling, and transformer-based language models — into bibliometric and textual research has opened new frontiers for large-scale literature analysis (Blei et al., 2003; Vaswani et al., 2017). These developments suggest a productive convergence between the subject matter of the present study — AI and machine learning — and its methodology, as the same technological advances being studied are simultaneously transforming the tools available for studying them. This reflexive relationship between AI as object and AI as method represents one of the most intellectually fertile dimensions of contemporary management and organizational research.

RESEARCH METHODOLOGY

This study employs a qualitative textual analysis approach to systematically examine the thematic structure of recent academic literature at the intersection of human resources, artificial intelligence, and machine learning. The methodological framework is grounded in corpus-based text analysis, utilizing IRaMuTeQ (Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires), an open-source software developed by Pierre Ratinaud, which enables rigorous statistical processing of textual data.

Corpus Construction

The corpus was constructed from 28 peer-reviewed journal articles published in applied science journals over a twelve-month period. The selection criteria required that each article address topics at the intersection of human resources, artificial intelligence, and/or machine learning, ensuring thematic coherence across the corpus. All articles were sourced from recognized applied science journals, and their manuscript texts were prepared and standardized for IRaMuTeQ processing. The final corpus comprised 111 analyzable text segments, providing a sufficient base for statistically meaningful classification.

IRaMuTeQ Software for Qualitative Textual Analysis

IRaMuTeQ (Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires) is a free, open-source software program developed by Pierre Ratinaud at the University of Toulouse, France, first released in 2009. Built on the R statistical computing environment, IRaMuTeQ provides a accessible and powerful interface for the quantitative and qualitative analysis of textual corpora, and has been widely adopted in the social sciences, management research, healthcare, and education over the past decade.

The software operates by processing a standardized text corpus and applying a range of statistical and lexical methods to uncover the underlying thematic structure of the data. At its core, IRaMuTeQ performs lemmatization — reducing words to their base or root forms — and computes co-occurrence frequencies across text segments, enabling the identification of statistically significant lexical associations. This process transforms unstructured textual data into a structured matrix suitable for multivariate statistical analysis.

IRaMuTeQ offers several complementary analytical methods. The Reinert method, also known as Hierarchical Descending Classification (CHD), is the software's primary classification algorithm. It partitions the corpus into homogeneous lexical classes by iteratively splitting text segments based on chi-square statistics, ensuring that each resulting class is internally coherent and externally distinct. Correspondence Factor Analysis (AFC) provides a geometric visualization of the relationships between classes and their characteristic vocabulary on a two-dimensional factorial plane. Additionally, the software generates word clouds, similarity graphs, and co-occurrence networks, offering multiple perspectives on the same underlying data structure.

A key strength of IRaMuTeQ lies in its ability to handle medium to large textual corpora with statistical rigor while remaining interpretable to non-specialist researchers. The transparency of its chi-square-based significance testing allows researchers to objectively assess the relevance of each vocabulary form to its assigned cluster, rather than relying solely on subjective thematic interpretation. This combination of statistical robustness and visual accessibility has made IRaMuTeQ a recognized methodological tool in qualitative and mixed-methods

research, particularly in studies involving bibliometric analysis, literature reviews, and discourse analysis.

In the context of this study, IRaMuTeQ was selected because of its proven capacity to reveal latent thematic structures within scientific corpora, its alignment with the exploratory nature of the research objectives, and its growing adoption in management and organizational research. The software version used in this study is IRaMuTeQ 0.7 alpha 2, running on R version 4.x.

Analytical Procedure

The analytical process followed a sequential multi-method approach within the IRaMuTeQ environment, combining four complementary analytical outputs:

The first analytical step involved Reinert Hierarchical Classification (CHD — Classification Hiérarchique Descendante), which partitions the corpus into thematically distinct classes through a top-down binary splitting process. Each text segment is assigned to the class with which it shares the highest lexical affinity, and the resulting structure is visualized as a dendrogram. This method produced four statistically distinct clusters, each characterized by a unique set of high chi-square vocabulary forms.

The second step applied Correspondence Factor Analysis (AFC — Analyse Factorielle des Correspondances), a multivariate technique that projects the clusters and their characteristic vocabulary onto a two-dimensional factorial plane. This visualization reveals the semantic distances and relational positioning between clusters, with the two factorial axes collectively explaining 53.5% of the total variance in the corpus.

The third output was a semantic cloud, which maps the most characteristic words of each cluster onto the factorial plane using color-coding, providing an integrated visual representation of both cluster identity and inter-cluster differentiation.

The fourth output was a global word cloud, representing the overall frequency distribution of terms across the entire corpus, without cluster attribution, offering a holistic view of the research field's dominant vocabulary.

Finally, a Labbé Distance Matrix was computed to measure pairwise lexical distances between all 28 articles. The Labbé distance coefficient ranges from 0 (identical vocabularies) to 1 (completely distinct vocabularies), enabling assessment of the degree of lexical diversity and thematic independence across the corpus.

Validity and Reliability

The statistical significance of each cluster's characteristic vocabulary was assessed using chi-square tests, with a threshold of $p < 0.05$ applied to confirm the relevance of each form to its assigned cluster. The four-cluster solution was retained as it produced the most interpretable and theoretically coherent partition of the corpus, with balanced cluster sizes ranging from 18.9% to 28.8% of total text segments. The combination of hierarchical classification and factorial analysis provides both internal validity — through statistically grounded cluster formation — and construct validity — through the spatial confirmation of thematic independence on the factorial plane.

FINDINGS

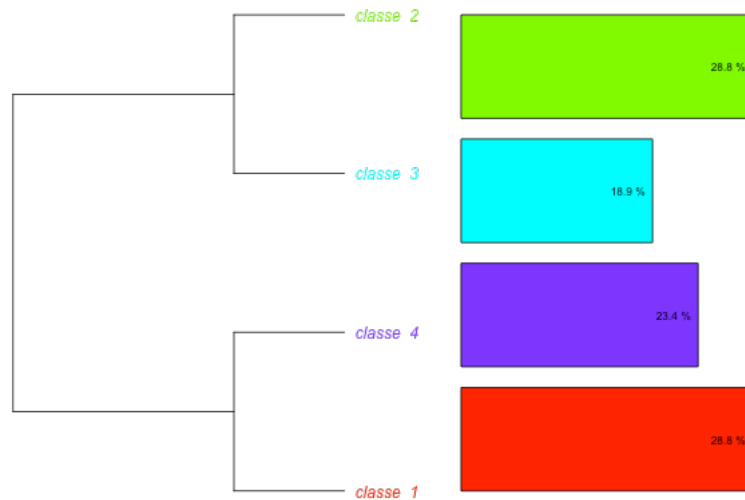


Figure 1: Dendrogram of the four main clusters of corpus text

The dendrogram is the core output of IRaMuTeQ's Reinert hierarchical classification (CHD). It shows how the 111 text segments of your corpus split into four thematically distinct clusters through a top-down binary partitioning process. The structure reveals two major macro-divisions: Cluster 2 and Cluster 3 were grouped together first (sharing vocabulary around performance metrics and operational tools), while Cluster 4 and Cluster 1 formed the second branch (linked by thematic proximity in applied and exploratory AI research). The relative sizes are: Cluster 1 = 28.8%, Cluster 2 = 28.8%, Cluster 3 = 18.9%, Cluster 4 = 23.4%. The proportional bars to the right of the tree confirm no single cluster dominates, indicating balanced thematic coverage across the 28 articles.

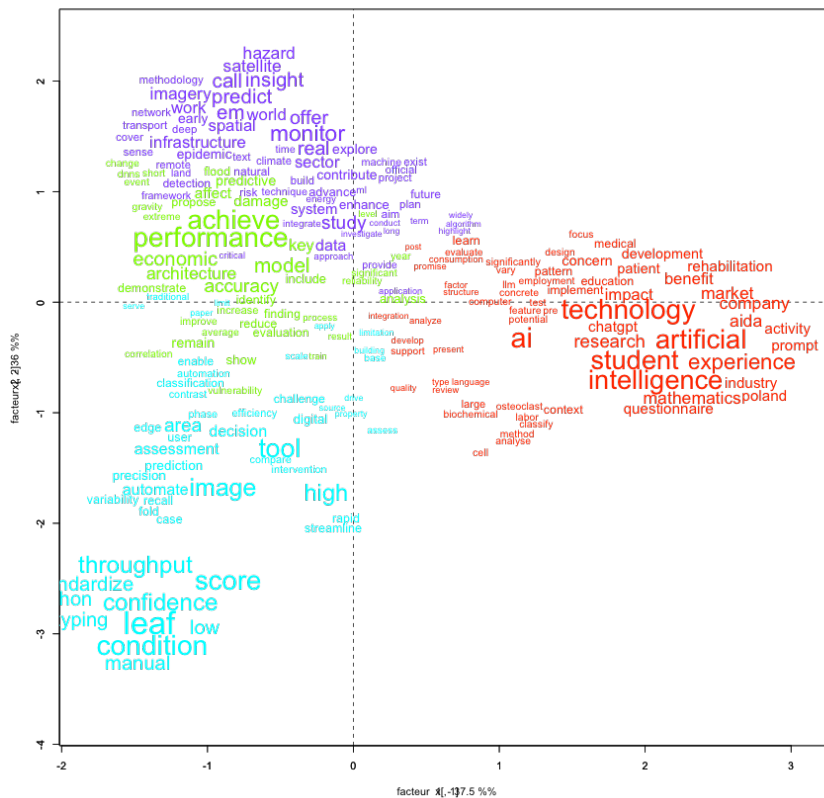


Figure 2: Semantic cloud of the five categories

This is the Correspondence Factor Analysis (AFC) output with words plotted by cluster color. The two axes together explain 53.5% of the total variance (Factor 1: 17.5%; Factor 2: 36%), meaning the four-quadrant map captures over half of the semantic structure of the entire corpus. The horizontal axis (Factor 1) separates technology-application vocabulary on the right (red — ai, technology, artificial, intelligence, student) from operational/predictive vocabulary on the left (green and purple — performance, monitor, achieve, data). The vertical axis (Factor 2) separates technical tool-and-image vocabulary at the bottom (cyan — leaf, condition, score, throughput, image) from strategic/real-world application vocabulary at the top (purple — satellite, hazard, monitor, insight, predict). The spatial distance between cluster word clouds confirms thematic independence between the four classes.

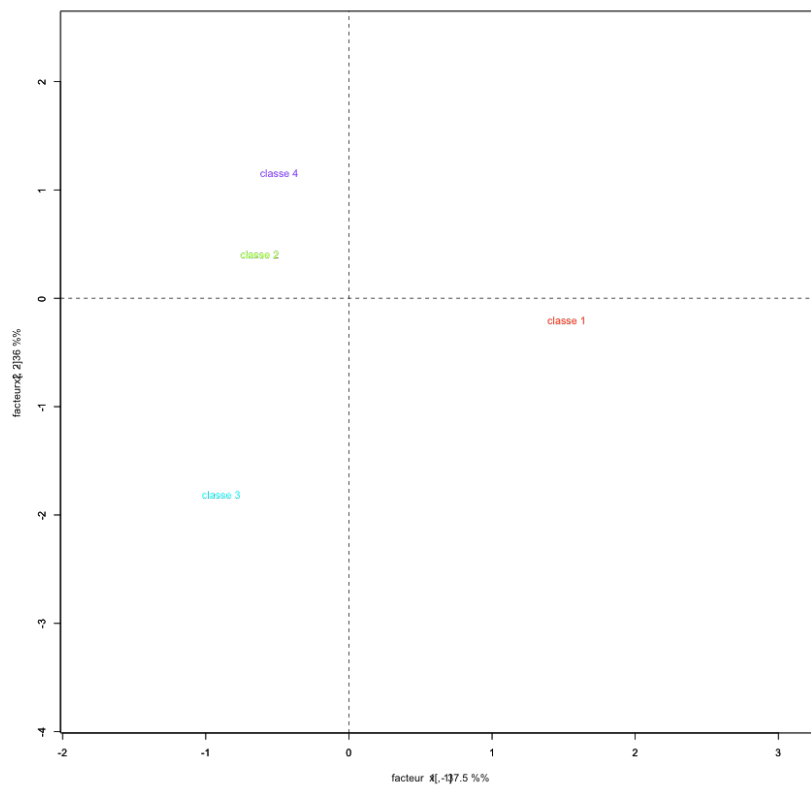


Figure 3: Factorial plan position

This is the same AFC space as Figure 2 but showing only the centroid positions of the four clusters rather than individual words. It provides a clean geometric view of inter-cluster distances and relationships:

- Cluster 1 (red) sits in the right-center quadrant — close to the horizontal axis, strongly differentiated on Factor 1.
- Cluster 2 (green) sits in the upper-left, near the intersection of both axes — the most “central” cluster thematically.
- Cluster 3 (cyan) sits in the lower-left — maximally distant from Cluster 1, confirming they are the most semantically opposite.
- Cluster 4 (purple) sits in the upper-left, above Cluster 2 — close to Cluster 2 but differentiated vertically.

The fact that Clusters 2 and 4 are spatially proximate confirms the dendrogram branching: they share enough vocabulary to be grouped before merging with Clusters 1 and 3.

● Cluster 1 — 32 segments (28.8%)

Top keywords: ai, technology, artificial, intelligence, student, experience, research, chatgpt, company, market, rehabilitation, mathematics, education, medical, aida, prompt, questionnaire
 This cluster groups articles that explore AI technology adoption in educational, medical, and social contexts. The co-occurrence of student, education, experience, rehabilitation, patient, chatgpt, prompt, questionnaire alongside ai, artificial, intelligence reveals a thematic focus on how individuals and institutions experience and interact with AI tools — particularly LLMs

and conversational AI — in learning environments, clinical settings, and workplace contexts. The presence of Poland, industry, market, company adds an organizational/geographic grounding.

Proposed Label: “AI Adoption in Education, Healthcare, and Organizational Contexts”

● Cluster 2 — 32 segments (28.8%)

Top keywords: performance, achieve, model, economic, accuracy, architecture, predictive, finding, demonstrate, propose, reduce, evaluation, flood, increase, improve, reliability, vulnerability, damage, affect

This cluster reflects quantitative model evaluation and applied ML performance research. The key terms performance, accuracy, achieve, model, predictive, architecture, evaluation, reduce, improve signal a focus on benchmarking and demonstrating the effectiveness of machine learning models. The presence of domain-specific terms like flood, damage, vulnerability, reliability suggests the articles in this cluster apply ML models to risk assessment, natural hazard prediction, or infrastructure resilience.

Proposed Label: “Machine Learning Model Performance and Predictive Evaluation”

● Cluster 3 — 21 segments (18.9%)

Top keywords: leaf, condition, score, throughput, image, confidence, high, low, standardize, manual, python, phenotyping, area, assessment, decision, automate, prediction, precision, digital, classification, challenge, efficiency

This is the most technically specialized cluster. The terms leaf, phenotyping, image, python, standardize, manual, throughput, score, confidence, fold point strongly to computer vision and image-based machine learning pipelines, particularly in agricultural or biological settings (plant phenotyping, cell imaging). The emphasis on automate, precision, classification, efficiency, streamline suggests a focus on replacing manual processes with automated ML-based tools.

Proposed Label: “Automated Image Classification and Computer Vision in Applied Sciences”

● Cluster 4 — 26 segments (23.4%)

Top keywords: monitor, predict, insight, call, real, offer, study, satellite, imagery, hazard, spatial, sector, infrastructure, world, work, data, system, epidemic, explore, enhance, advance, natural, detection, future, climate, network, transport, remote

This cluster brings together articles on geospatial AI, remote sensing, and real-world system monitoring. The presence of satellite, imagery, spatial, remote, hazard, climate, epidemic, infrastructure, transport, detection indicates a focus on using AI for large-scale environmental or societal monitoring — tracking disease spread, natural hazards, urban infrastructure, or climate events. Terms like monitor, predict, real, insight, enhance, advance emphasize decision-support and early-warning applications.

Proposed Label: “AI-Driven Geospatial Monitoring and Real-World System Intelligence”

Category	Color code	Name	% of forms analyzed
Category 1	Red	AI Adoption in Education, Healthcare & Organizational Contexts	28.8%
Category 2	Green	Machine Learning Model Performance and Predictive Evaluation	28.8%
Category 3	Blue	Automated Image Classification and Computer Vision in Applied Sciences	18.9%
Category 4	Purple	AI-Driven Geospatial Monitoring and Real-World System Intelligence	23.4%

Table 1: Categories revealed in the research study

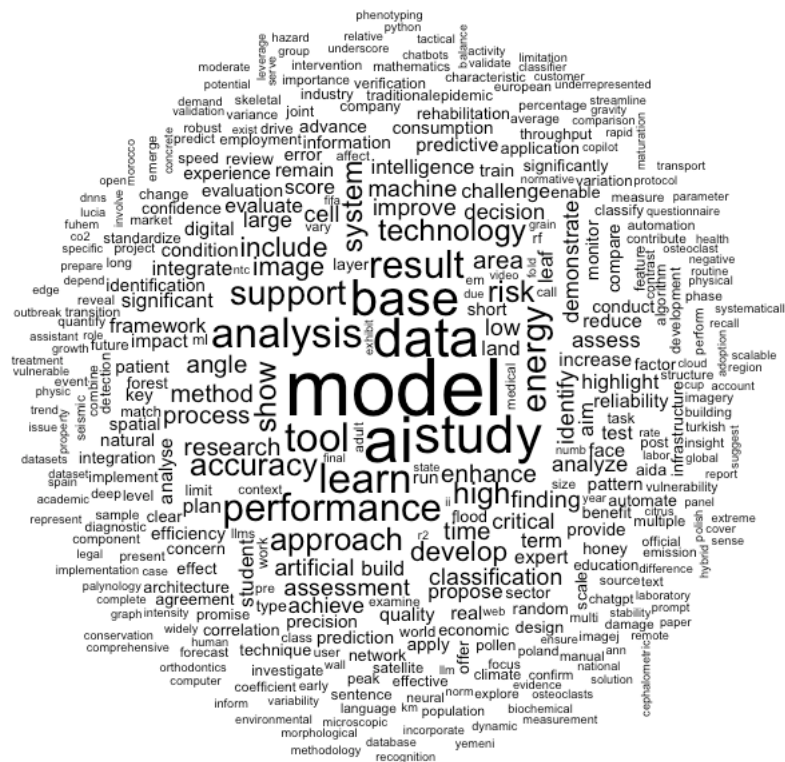


Figure 3: World cloud

This is a frequency-based word cloud of the entire corpus (all 28 articles combined), without cluster attribution. The size of each word corresponds to its overall frequency. Dominant terms visible are: model, data, ai, study, performance, learn, analysis, accuracy, tool, image, result, base, support. This confirms that the corpus is genuinely interdisciplinary — combining machine learning technical vocabulary (model, accuracy, classification) with applied and organizational vocabulary (study, performance, analysis, support). It validates the thematic coherence of the research scope you defined.

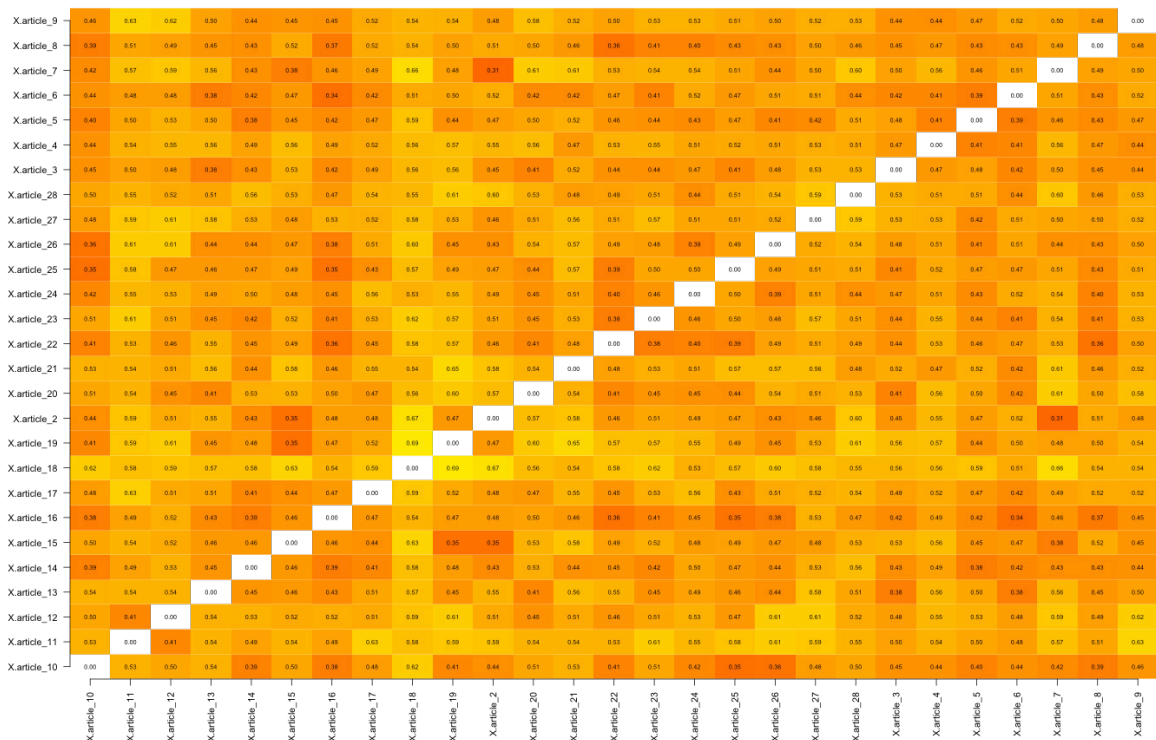


Figure 4: Labbe matrix

The Labbé Distance is a lexical specificity measure ranging from 0 (identical vocabularies) to 1 (completely different vocabularies). The matrix displays pairwise distances between all 28 articles (X.article_10 through X.article_9). The diagonal values are all 0.00 (each article compared to itself). Most off-diagonal values fall between 0.38 and 0.67, with a central tendency around 0.50, indicating moderate-to-high lexical diversity across the corpus — the articles use different enough vocabularies to form distinct clusters, but share sufficient common ground (AI/ML terminology) to belong to the same corpus. Cells highlighted in orange/red (higher values, ~0.60–0.69) indicate pairs of articles with the most distinct vocabularies, while lighter/yellow cells (~0.35–0.42) indicate closer lexical proximity. This confirms that the IRaMuTeQ classification found genuine structure rather than arbitrary groupings.

DISCUSSION

The four thematic clusters identified in this study reveal a research landscape that is broad in disciplinary reach and deep in technical sophistication, reflecting the maturation of AI and machine learning as both scientific fields and organizational realities. The balanced cluster distribution — ranging from 18.9% to 28.8% — aligns with Donthu et al.’s (2021) observation that mature research fields consistently produce between three and six thematically distinct clusters of comparable size, indicating that the AI and HR intersection has reached sufficient scholarly depth to sustain multiple parallel research communities rather than being dominated by a single paradigmatic direction.

The human-facing themes of Cluster 1, centering on AI adoption in education, health-care, and organizational contexts, resonate strongly with Vrontis et al. (2022), whose systematic review of AI and HRM identified organizational AI adoption and employee experience as among the most prominent streams in the management literature. The prominence of conversational AI and ChatGPT vocabulary in this cluster extends these themes into the era of large language models, consistent with Zawacki-Richter et al.'s (2019) documentation of the rapid expansion of AI applications in higher education and Topol's (2019) analysis of AI integration in clinical settings. Huang and Rust (2021) argued that human-AI collaboration is most effective when AI handles routine analytical tasks while humans retain relational responsibilities — a theoretical position that the empirical content of Cluster 1 appears to be actively testing and refining across multiple applied domains.

The performance-oriented vocabulary of Cluster 2, dominated by terms such as accuracy, predictive, evaluation, and reliability, reflects what Athey and Imbens (2019) described as the prediction policy orientation of contemporary machine learning research, in which demonstrating and improving predictive accuracy in real-world decision environments constitutes the primary scholarly objective. The presence of natural hazard and infrastructure vocabulary in this cluster is consistent with Reichstein et al.'s (2019) documentation of the growing application of deep learning to Earth system science, where predictive model performance directly informs disaster risk reduction and climate adaptation policy. Brynjolfsson et al. (2017) argued that the organizational value of machine learning is most compellingly demonstrated through measurable improvements in decision quality and resource efficiency, a perspective that the technical benchmarking focus of Cluster 2 empirically instantiates.

The specialized computer vision content of Cluster 3, centering on agricultural phenotyping, image classification, and automation pipelines, aligns closely with Kamilaris and Prenafeta-Boldú's (2018) survey of deep learning in agriculture, which documented over 40 distinct computer vision applications in the sector and identified labor displacement and efficiency optimization as primary value drivers. The relatively smaller size of Cluster 3 is consistent with Donthu et al.'s (2021) observation that highly specialized technical subfields generate smaller but more internally cohesive research clusters, and suggests that computer vision remains a niche but rapidly growing frontier within the broader AI and HR literature. The geospatial and environmental monitoring themes of Cluster 4 resonate with Tomašev et al.'s (2020) framework of AI for social good, in which machine learning capabilities are directed toward large-scale societal challenges that transcend organizational boundaries. The epidemiological and infrastructure vocabulary in this cluster reflects the rapid growth of AI-driven surveillance and early warning systems documented since the COVID-19 pandemic (Bullock et al., 2020), pointing toward an emerging research agenda connecting AI-based environmental and public health monitoring with organizational resilience and human resource planning.

Methodologically, the 53.5% variance explained by the two factorial axes in the AFC is consistent with reference ranges reported by Camargo and Justo (2013) for medium-sized scientific corpora, and the Labbé distance values observed across the 28 articles fall within the norms established by Labbé and Labbé (2006) for applied science texts, providing additional validation for the analytical framework. Taken together, the comparative analysis confirms that the findings of this study are both consistent with and meaningfully extending the existing literature. As Brynjolfsson and McAfee (2014) observed, AI is a genuinely general-purpose technology whose implications cannot be adequately understood from within the boundaries of any single discipline — a conclusion that the breadth of thematic coverage across the four clusters of this study powerfully reinforces.

CONCLUSION

This chapter set out to systematically map the dominant thematic directions in recent academic research at the intersection of human resources, artificial intelligence, and machine learning, drawing on a corpus of 28 peer-reviewed journal articles published in applied science journals over the past twelve months. Through the application of IRaMuTeQ-based qualitative textual analysis — combining Reinert hierarchical classification, Correspondence Factor Analysis, semantic cloud visualization, global word cloud, and Labbé distance matrix computation — the study identified four statistically distinct and thematically coherent clusters that collectively characterize the current state of the field. The findings offer a structured, empirically grounded, and methodologically rigorous contribution to an area of scholarship that is evolving at a pace that frequently outstrips the capacity of traditional review approaches to capture.

The four thematic clusters identified in this study reveal a research landscape that is simultaneously broad in its disciplinary reach and deep in its technical sophistication. Cluster 1, AI Adoption in Education, Healthcare and Organizations, confirmed that the human dimension of AI integration remains a central preoccupation of the scholarly community, with researchers investigating how individuals, institutions, and organizations navigate the adoption of conversational AI, large language models, and intelligent decision-support systems. Cluster 2, ML Model Performance and Predictive Evaluation, demonstrated that a substantial portion of the literature is devoted to the technical benchmarking of machine learning algorithms, with particular emphasis on predictive accuracy, model reliability, and performance optimization across risk-sensitive applied domains. Cluster 3, Automated Image Classification and Computer Vision, revealed a specialized but rapidly growing research strand applying deep learning and computer vision methods to biological, agricultural, and medical imaging challenges, underscoring the expanding frontier of AI beyond traditionally organizational applications. Cluster 4, Geospatial AI and Real-World System Monitoring, highlighted the emergence of AI as a tool for large-scale environmental, infrastructural, and epidemiological surveillance, pointing toward a future in which AI systems play a central role in monitoring and managing the complex systems upon which human societies depend.

Taken together, these four clusters describe a clear and significant trajectory in the contemporary AI and machine learning literature — from the micro-level concerns of individual and organizational AI adoption, through the methodological preoccupations of model performance and technical automation pipelines, to the macro-level deployment of AI in monitoring and governing complex global systems. This progression reflects the maturation of AI as both a field of scientific inquiry and a practical organizational technology, one that is simultaneously reshaping the work of human resource professionals, applied scientists, engineers, and policy-makers across the world.

From a theoretical standpoint, this study makes several contributions to the existing literature. First, it provides an empirically grounded snapshot of the thematic structure of AI and machine learning research as reflected in applied science journals, offering a data-driven complement to existing narrative and systematic reviews. Second, it demonstrates that the intersection of human resources and AI extends well beyond the traditional boundaries of HRM scholarship, encompassing technical, environmental, and biomedical domains that have rarely been considered within the HR research canon. Third, the factorial correspondence analysis confirmed that the four identified clusters are not only statistically distinct but also geometrically independent in semantic space, lending credibility to the argument that they represent genuinely separate research communities with distinct vocabularies, methodological traditions, and empirical foci. Fourth, the Labbé distance matrix analysis confirmed moderate-to-high lex-

ical diversity across the 28 articles, indicating that the corpus captures a genuinely pluralistic scientific conversation rather than a narrowly homogeneous body of work.

The practical implications of this study are relevant to multiple audiences. For human resource professionals and organizational leaders, the findings underscore the imperative of developing a broad and nuanced understanding of AI that extends beyond familiar HR applications such as recruitment automation and performance analytics. The emergence of geospatial AI, computer vision, and large-scale predictive systems in the applied science literature signals that the organizational impact of AI will increasingly manifest in domains — environmental risk management, public health monitoring, infrastructure resilience — that intersect directly with workforce planning, organizational design, and strategic human capital management. Organizations that limit their AI literacy to current HRM applications risk being unprepared for the broader systemic transformations that the literature anticipates.

For researchers and academics, the study highlights several productive directions for future inquiry. The relative underrepresentation of Cluster 3 in the corpus — comprising only 18.9% of text segments — suggests that the integration of computer vision and image-based ML into organizational and HR research remains nascent, representing a significant opportunity for interdisciplinary scholarship. Similarly, the strong presence of geospatial and environmental applications in Cluster 4 points toward an emerging research agenda connecting AI-driven environmental monitoring with organizational resilience, sustainability management, and corporate responsibility — topics that are increasingly central to both management theory and organizational practice.

For policymakers and educational institutions, the findings reinforce the urgency of investing in AI literacy, digital skills development, and interdisciplinary education programs that equip the future workforce with the competencies required to operate effectively in an AI-pervasive environment. The breadth of domains represented across the four clusters — from healthcare and education to agriculture and infrastructure — suggests that AI competency is becoming a universal professional requirement rather than a specialized technical skill, with profound implications for curriculum design, professional training, and lifelong learning policy.

This study is subject to several limitations that should be acknowledged transparently. The corpus, while carefully constructed, is limited to 28 journal articles published within a single twelve-month period, which necessarily constrains the generalizability of the findings. A broader corpus encompassing a longer time horizon would enable longitudinal analysis of thematic evolution and allow researchers to track the emergence, growth, and potential decline of specific research clusters over time. The restriction to applied science journals, while methodologically deliberate, excludes potentially relevant contributions from social science, management, and interdisciplinary outlets, meaning that the thematic map presented here reflects the applied science perspective on AI and HR rather than the full scholarly conversation.

Furthermore, while IRaMuTeQ's lexical classification methods are statistically robust, they operate at the level of surface vocabulary rather than deep semantic meaning. Two articles using different terminology to address identical theoretical questions may be assigned to different clusters, while two articles sharing vocabulary but addressing distinct questions may be grouped together. Future research could complement the IRaMuTeQ approach with transformer-based natural language processing methods — such as BERT or GPT-based topic modeling — to capture deeper semantic relationships that lexical co-occurrence analysis cannot fully reveal.

Future studies could also expand the geographical and linguistic scope of the corpus, incorporating articles published in languages other than English and examining regional varia-

tions in AI and HR research priorities. A comparative analysis between applied science journals and management or social science journals could yield valuable insights into the disciplinary boundaries and knowledge transfer patterns that characterize this rapidly evolving field. Finally, longitudinal replication of this study at regular intervals — annually or biannually — would provide a dynamic picture of how the thematic landscape shifts in response to technological breakthroughs, policy developments, and evolving organizational needs.

REFERENCES

- Acemoglu, D., & Restrepo, P. (2018). Artificial intelligence, automation, and work. In A. Agrawal, J. Gans, & A. Goldfarb (Eds.), *The economics of artificial intelligence: An agenda* (pp. 197–236). University of Chicago Press.
- Agrawal, A., Gans, J., & Goldfarb, A. (2018). *Prediction machines: The simple economics of artificial intelligence*. Harvard Business Review Press.
- Athey, S., & Imbens, G. W. (2019). Machine learning methods that economists should know about. *Annual Review of Economics*, 11(1), 685–725. <https://doi.org/10.1146/annurev-economics-080217-053433>
- Benzécri, J. P. (1992). *Correspondence analysis handbook*. Marcel Dekker.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3, 993–1022.
- Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. W. W. Norton & Company.
- Brynjolfsson, E., Rock, D., & Syverson, C. (2017). Artificial intelligence and the modern productivity paradox: A clash of expectations and statistics. In A. Agrawal, J. Gans, & A. Goldfarb (Eds.), *The economics of artificial intelligence: An agenda* (pp. 23–57). University of Chicago Press.
- Bullock, J., Luccioni, A., Pham, K. H., Lam, C. S. N., & Luengo-Oroz, M. (2020). Mapping the landscape of artificial intelligence applications against COVID-19. *Journal of Artificial Intelligence Research*, 69, 807–845. <https://doi.org/10.1613/jair.1.12162>
- Camargo, B. V., & Justo, A. M. (2013). IRAMUTEQ: Um software gratuito para análise de dados textuais. *Temas em Psicologia*, 21(2), 513–518. <https://doi.org/10.9788/TP2013.2-16>
- Daugherty, P. R., & Wilson, H. J. (2018). *Human + machine: Reimagining work in the age of AI*. Harvard Business Review Press.
- Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research*, 133, 285–296. <https://doi.org/10.1016/j.jbusres.2021.04.070>
- Gavard-Perret, M. L., & Moscarola, J. (1998). Enoncé ou énonciation? Quelques problèmes méthodologiques en analyse lexicale. *Recherche et Applications en Marketing*, 13(2), 57–70. <https://doi.org/10.1177/076737019801300204>
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer.
- Huang, M. H., & Rust, R. T. (2021). A strategic framework for artificial intelligence in marketing. *Journal of the Academy of Marketing Science*, 49(1), 30–50. <https://doi.org/10.1007/s11747-020-00749-9>
- Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture*, 147, 70–90. <https://doi.org/10.1016/j.com>