 ISSN 2776-8007

K-Means Clustering for Segmenting AI Survey Respondents: Analysis of Information Sources and Impact Perceptions

Evelyn[1,*], Satrio Pradono Suryodiningrat[2], Masmur Tarigan[3]

[1] School of Business Management, Petra Christian University, Indonesia

[2]School of Computing and Creative Arts, University of Bina Nusantara, Jakarta, Indonesia

[3]Faculty of Computer Science, Esa Unggul University

(Received: October 1, 2024; Revised: November 30, 2024; Accepted: February 1, 2025; Available online: April 1, 2025)

Abstract

This study employs K-Means clustering to analyze survey data from 91 university students, aiming to segment respondents based on their information-seeking behaviors (Question 2) and impact perceptions (Question 3) of artificial intelligence (AI). Two distinct clusters emerged: Optimistic Problem Solvers, who favor formal channels such as scholarly websites, peer-reviewed papers, and guided discussions, and express strong confidence in AI's problem-solving capabilities with low concern for job displacement or dehumanization; and Critical Watchers, who rely more on informal, rapidly updated media (e.g., social platforms, general web searches) and exhibit heightened apprehension regarding AI's socio-economic and ethical risks. Demographically, the former group skews toward sophomores with consistent GPAs and quantitatively oriented majors, while the latter displays broader disciplinary representation, balanced gender composition, and greater academic variability. These findings validate a dual-dimensional segmentation framework that integrates source behavior with perceptual orientation, highlighting the inadequacy of one-size-fits-all AI education. The study recommends differentiated instructional strategies, deep-dive, research-oriented modules for problem-solvers and trust-building, narrative-driven outreach for watchers, and outlines future research directions including larger, multi-institutional samples, longitudinal tracking, and mixed-methods approaches to refine and validate these profiles.

Keywords: K-Means clustering, AI perception, Information sources, Student segmentation, Higher education

1. Introduction

The rapid proliferation of Artificial Intelligence (AI) technologies has ushered in a transformative era for higher education, fundamentally reshaping the ways in which university students engage with learning materials, collaborate with peers, and develop the critical skills required for tomorrow's workforce. As institutions worldwide invest in the digital infrastructures necessary to support AI-driven tools ranging from adaptive learning platforms to intelligent tutoring systems research by Chen et al. [1] and Chan and Hu [2] demonstrates that campuses with robust technological ecosystems report significantly higher levels of student engagement in self-directed learning activities, underscoring the pivotal role that institutional readiness plays in maximizing the pedagogical benefits of AI integration. At the same time, studies by Quinde et al. [3] highlight that students' technological literacy encompassing competencies such as data fluency, algorithmic reasoning, and ethical awareness is closely correlated with employability outcomes, as employers increasingly prioritize graduates who can critically assess and apply AI capabilities to real-world problems.

However, the successful incorporation of AI into curricula depends not only on the physical and digital infrastructure but also on the cultivation of a campus culture that values innovation, experimentation, and responsible use of emerging technologies, Jo and Bang [4] argue that fostering such a culture requires proactive professional development for faculty, interdisciplinary collaboration among academic units, and institutional policies that incentivize the ethical deployment of AI in teaching and research. Despite these promising trends, the integration of AI into higher education is not without its challenges: ethical considerations such as algorithmic bias, data privacy concerns, and equitable

*Corresponding author: Evelyn (evelyn@petra.ac.id)
DOI: https://doi.org/10.47738/ijaim.v5i1.99

access to AI resources remain front and center in the scholarly discourse. Without rigorous oversight and transparent governance frameworks, AI systems risk exacerbating existing educational inequalities, as under-resourced students may lack access to high-quality internet connectivity, hardware, or training necessary to leverage AI tools effectively.

Concurrently, evidence from Sarwari and Adnan [5] and Sun and Zhou [6] suggests that AI applications such as large-language chatbots (e.g., ChatGPT) can significantly enhance critical thinking, facilitate rapid prototyping of academic assignments, and support iterative feedback loops between students and instructors but only when accompanied by well-designed pedagogical scaffolds and clear guidelines addressing responsible usage. Mustopa et al. [7] emphasize that, as AI continues to evolve, educational institutions must remain agile in updating curricula, integrating hands-on AI modules, and embedding ethics-in-AI conversations throughout academic programs to ensure that students not only master technical skills but also develop the judgment and empathy essential for socially responsible innovation.

Central to these strategic efforts is a deep understanding of student user profiles, that is, nuanced insights into learners' prior knowledge of AI, preferred channels for obtaining information, attitudinal predispositions toward AI benefits and risks, and demographic factors that shape technology adoption. Kaur et al. [8] introduce an intelligent profiling system that synthesizes personalized data such as academic transcripts, professional experiences, and digital footprints to recommend tailored learning pathways, while Kordahi [9] explores digital identity frameworks designed to track user interactions with online resources and dynamically adapt content recommendations. Complementing these approaches, Blanco et al. [10] demonstrate how applying user-centered design principles, through the creation of personas and scenario-driven interfaces can optimize navigation experiences and boost engagement with AI tools across diverse student segments. Moreover, research by Kraft and Bolves [11] illuminates how socioeconomic background influences family and peer support structures in engaging with educational technology, reinforcing the notion that equity-minded profiling is essential for crafting interventions that resonate with learners from varied contexts.

Yet, despite these advances, there remains a conspicuous gap in the literature concerning the segmentation of AI survey respondents by the very dimensions that most directly inform instructional design and policy: namely, the information sources students rely upon (e.g., websites, academic papers, social media, peer discussions) and the multifaceted perceptions they hold about AI's impacts (ranging from its problem-solving potential to concerns about job displacement, dehumanization, and societal governance). Qaladi et al. [12] caution that traditional self-reported surveys often fail to capture the longitudinal dynamics of attitude formation, thereby limiting the ability of educators and policymakers to design sustained, targeted campaigns. In professional contexts, Valerio [13] alongside Chan and Hu [2] illustrate how attitudinal clusters can predict differential adoption behaviors in educational settings.

To bridge this critical gap, the present study applies K-Means clustering a widely used unsupervised learning technique to segment a cohort of 91 university students who responded to a structured AI survey. By transforming multi-response items (Q2) into binary indicators of information-source usage and similarly encoding perception items (Q3) into discrete attributes, we construct a comprehensive feature matrix that captures both behavioral and attitudinal dimensions. The optimal number of clusters is determined through the Elbow Method, examining inertia reductions across $k = 2$ to 6 and validated via the Silhouette Score, ensuring that clusters are both compact and well separated. Subsequent visualizations employ principal component analysis (PCA) and t-distributed stochastic neighbor embedding (t-SNE) to project high-dimensional data into two-dimensional space, while heatmaps, stacked bar charts, and radar (spider) charts enable intuitive interpretation of cluster centroids across Q2 and Q3 variables. We further explore how demographic factors gender, year of study, academic major, number of exams passed, and GPA correlate with cluster membership, thereby illuminating the intersections between personal background and AI information behavior.

Our analysis reveals two distinct respondent segments: the first, labeled Optimistic Problem Solvers, is characterized by a strong proclivity toward leveraging AI for complex problem-solving tasks and a higher reliance on academic sources such as books and peer discussions; the second, termed Critical Watchers, exhibits elevated concerns about AI's potential to displace human labor, erode humanistic values, and exert undue influence over societal norms, often complemented by a greater dependence on social media and general-interest internet sources. These findings not only substantiate the theoretical linkages posited by foundational work on technology acceptance and media effects but also furnish practical guidance for higher education stakeholders seeking to tailor AI curricula and outreach initiatives to specific learner profiles.

By delineating the contours of AI information ecosystems and attitudinal landscapes, this study contributes methodologically to the field of educational data mining and practically to the design of equitable, effective AI-enhanced learning environments. Through targeted profiling and segmentation, universities can craft bespoke training modules, ethical guidelines, and communication campaigns that address the distinct needs and concerns of each student segment thereby maximizing engagement, fostering technological equity, and ultimately empowering learners to navigate the AI-mediated future with confidence and critical agency.

2. Literature Review

2.1 Clustering and Segmentation in Educational and Social Research

Clustering and segmentation methods have become indispensable tools in both social and educational research for decoding complex patterns of behavior and engagement. In the realm of social media analysis, Ma and Zhang [14] introduce a network-based segmentation framework that categorizes publics according to their interaction patterns and communication flows, demonstrating how insights into connectedness can reveal distinct community dynamics. This network paradigm carries over to educational contexts, Zhou [15] illustrates that clustering students by their online learning behaviors uncovers unique learner profiles, enabling instructors to tailor pedagogical strategies and personalize learning experiences through data-driven segmentation.

Beyond purely digital interactions, segmentation also informs targeted interventions across diverse domains. Kinnunen et al. [16] show how shared music experiences among Generation Z foster broader social identities and community cohesion an insight that parallels the benefits of collaborative learning environments, where peer networks significantly influence educational outcomes. Similarly, Mukoka et al. [17] demonstrate that interventions directed at specific community groups identified through social network characteristics maximize engagement and effectiveness in public health campaigns. Translating these approaches to academia, educators can leverage network-informed segmentation to identify and support at-risk or low-performing student cohorts, thereby enhancing overall learning outcomes.

While interdisciplinary connections to marketing and consumer behavior offer valuable methodological insights, some contexts such as Safeer's [18] study on online impulse buying in fashion retail do not directly inform clustering applications in education and were therefore omitted from this review. Collectively, the highlighted studies underscore the power of a networked approach to segmentation: by integrating social media analytics with education data mining, researchers and practitioners can develop comprehensive strategies that address the unique characteristics of distinct user groups and foster improved outcomes across both social and educational settings.

2.2 Applications of K-Means Clustering in Technology Adoption Studies

K-Means clustering has proven invaluable for segmenting users based on their interactions with new technologies across various domains. In the hospitality sector, Demirçiftçi et al. [19] applied K-Means to classify hotel guests into technology compassionates and casual users, demonstrating that differences in personality traits can significantly influence in-room technology adoption and usage patterns. Similarly, Sakalessy and Purnomo [20] used K-Means to evaluate IT department employees' performance, showing how clustering by technology utilization metrics can guide strategic workforce development and targeted training interventions.

In educational settings, K-Means has been leveraged to tailor instruction by grouping learners according to their digital competencies and engagement. Faisal et al. [21] segmented vocational high school students based on ICT skill levels, enabling educators to design differentiated learning pathways that address each group's specific needs. Qiao [22] extended this approach to civic education, mining big data from college curricula to identify patterns of student engagement with digital civic practices, thereby informing curriculum adjustments that enhance technology acceptance and participation.

Beyond sector-specific applications, K-Means also aids in understanding broader technology acceptance dynamics. Ibrahim et al. [23] incorporated the technology acceptance model into their clustering of AI adopters, revealing distinct adopter types whose characteristics can inform customized marketing and training strategies. Collectively, these studies underscore the versatility of K-Means clustering as a tool for uncovering actionable insights whether to optimize guest experiences, improve learning outcomes, refine organizational performance, or foster greater acceptance of emerging technologies.

2.3 Role of Information Sources and Impact Perceptions in AI Acceptance

The role of information sources in shaping user attitudes toward AI technologies is particularly pronounced across domains such as healthcare, education, and business. Li et al. [24] report that oncologists with stronger educational backgrounds place greater trust in AI tools. These findings underscore how well-informed users who perceive their information channels as trustworthy tend to view AI's benefits more favorably, thereby increasing their intention to adopt such technologies.

Perceived utility and ease of use, as outlined in the Technology Acceptance Model (TAM), further mediate AI acceptance. Edgington and Kasztelnik [25] demonstrate that systems judged to be both useful and user-friendly foster higher adoption rates, and Schepart et al. [26] emphasize that transparency and explainability in AI decision-making strengthen user trust, positive perceptions of AI can offset concerns about associated risks. In educational environments, Valerio [13] highlights that students' awareness of ethical considerations is crucial; addressing such concerns alongside technical capabilities is essential to promote widespread acceptance among both educators and learners.

Demographic and socio-professional factors also play a significant role in AI adoption. Jiang et al. [27] show that personal experience, professional background, and education level influence user receptivity, with individuals in tech-centric fields generally more open to AI tools than those in less digitalized sectors. In healthcare, Shinners et al. [28] find that practitioners in rural settings often with limited exposure to AI exhibit greater resistance compared to metropolitan counterparts. These insights point to the need for tailored information dissemination strategies that address varying levels of familiarity, professional contexts, and ethical concerns, thereby ensuring that AI adoption efforts resonate with the specific needs of diverse user groups.

2.4 Previous Studies on AI User Profiling in Academic Contexts

The exploration of user profiling in academic contexts, particularly regarding artificial intelligence (AI), provides essential insights into how students, educators, and administrators adapt to these technologies. Several studies have demonstrated that user profiling can enhance the effectiveness of AI applications in educational settings, ensuring that tools are customized to meet the specific needs and preferences of different users.

For instance, the work by Jiang et al. emphasizes that understanding the different acceptance attitudes and decision-making mechanisms of users interacting with AI is crucial. They categorize AI users into service providers and task substitutes, revealing how individual differences influence users' interactions with AI applications Jiang et al. [27]. This nuanced understanding of user profiling allows for the creation of AI tools that resonate more effectively with diverse user categories, thus promoting higher acceptance rates.

Moreover, Liu and Shi [29] discuss how user characteristics and AI attributes influence the intention to use AI systems. Their findings highlight the importance of understanding users' emotional and cognitive profiles, which can aid developers in designing AI technologies that align better with user expectations and enhance adoption [29]. Such considerations are particularly relevant in educational contexts, where AI tools are integrated into curricula and learning management systems.

In the domain of AI chatbots in academic libraries, understanding user interactions can optimize these tools for educational purposes [30]. The profiling of users based on their needs assists in identifying how AI can enhance academic research and resource navigation [30]. The study highlights concerns associated with the technology while proposing that tailored interactions can foster positive user experiences.

Furthermore, the study by Yu et al. [31] suggests that profiling can significantly improve AI outcomes in specialized settings, such as remote assistance frameworks. Personalized AI, which considers profiling elements like users' previous experiences and preferences, can enhance the training of AI models, ultimately leading to better service provision in educational environments.

Collectively, these studies underscore that user profiling is essential in understanding and enhancing the integration of AI technologies in educational contexts. Tailoring AI applications to meet the specific needs and preferences of users not only improves technology acceptance but also creates a more responsive and effective educational experience.

3. Methodology

The following diagram illustrates the steps involved in the data analysis process using the K-Means algorithm, as shown in figure 1.

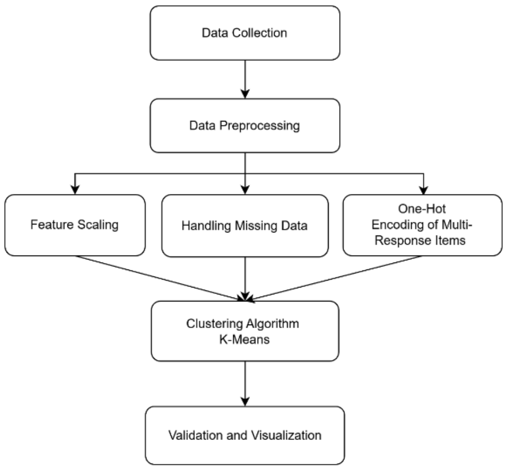

Figure 1. Research Methodology

3.1. Data Collection

This study employs a cross-sectional survey design to explore patterns in how university students access AI information and perceive its impacts. The dataset comprises 91 respondents who completed an online questionnaire distributed across multiple academic programs. Key survey items are summarized in table 1, and demographic variables are detailed in table 2.

Table 1. Overview of Key Survey Variables (Q2 & Q3 Items)

Variable Code	Description	Type
Q2#1.Internet	Accessed AI information via Internet (websites, blogs, search)	Binary
Q2#2.Books/Papers	Accessed AI information via books or scientific publications	Binary
Q2#3.Social_media	Accessed AI information via social media	Binary
Q2#4.Discussions	Obtained AI information through discussions (forums, peers, etc.)	Binary
Q2#5.NotInformed	Has not obtained any information about AI	Binary
Q3#1.AI_dehumanization	Perceives AI as potentially reducing human aspects	Binary
Q3#2.Job_replacement	Perceives AI as potentially replacing human jobs	Binary
Q3#3.Problem_solving	Perceives AI as potentially aiding problem solving	Binary
Q3#4.AI_ruling_society	Perceives AI as potentially governing societal order	Binary

Table 2. Overview of Demographic Variables

Variable Code	Description	Type
Q12.Gender	Respondent’s gender (Male / Female)	Categorical
Q13.Year_of_study	Year of study (1 = Freshman, 2 = Sophomore)	Ordinal
Q14.Major	Academic major/discipline (e.g., Computer Science)	Categorical
Q15.Passed_exams	Number of exams completed by the respondent	Numeric (count)
Q16.GPA	Cumulative Grade Point Average (0–10 scale)	Numeric (ratio)

Respondents were recruited via university mailing lists and social media groups over a two-week period. Participation was voluntary and anonymous, with informed consent obtained before survey submission. Data completeness checks ensured that only fully completed Q2 and Q3 responses were included in the final analysis.

3.2. Data Preprocessing

Before applying K-Means clustering, all survey responses were transformed into a suitable feature matrix. First, multi-response items in Q2 (information sources) and Q3 (impact perceptions) were converted into binary indicator variables via one-hot encoding, with each possible option represented by its own column indicating selection (1) or non-selection (0). Any respondents with incomplete answers for Q2 or Q3 were removed to maintain consistency in the clustering inputs, while sporadic missing values in demographic fields (such as GPA) were imputed, numeric gaps with the median and categorical gaps with the mode. Since the one-hot-encoded Q2 and Q3 features are already on a common binary scale, no further scaling was necessary for clustering; however, when demographic variables like GPA were later incorporated into supplementary analyses, they were standardized using z-score normalization to ensure comparability across different measurement scales.

3.3. Clustering Algorithm

K-means clustering works by minimizing the variance within clusters, ensuring that the items within a cluster are as similar as possible while being different from those in other clusters. The algorithm works iteratively by assigning data points to the nearest cluster centroid and recalculating the centroids until convergence. The formula used to calculate the distance between a data point x_i and a cluster centroid C_k is typically the Euclidean distance, given by:

$$d(x_i, C_k) = \sqrt{\sum_{j=1}^{n} (x_{ij} - c_{kj})^2} \tag{1}$$

The K-Means algorithm seeks to partition a set of n observations $\{x_1, x_2, \dots, x_n)$ into k clusters $C_1, C_2, \dots, C_k$ by minimizing the total within-cluster sum of squared distances (inertia), given by

$$J = \sum_{j=1}^{K} \sum_{x_i \in C_j} ||x_i - \mu_j||^2 \tag{2}$$

Each cluster centroid μ_j is defined as the mean of its members, $\mu_j = \frac{1}{|C_j|} \sum_{x_i \in C_j} x_i$. After initializing centroids (here via k-means++ to improve convergence stability), the algorithm alternates between assigning each point to the nearest centroid (by Euclidean distance) and recomputing centroids until assignments no longer change. To determine the optimal $k,$ the Elbow Method plots $J(k)$ for a range of k values and identifies the point of diminishing returns, while the Silhouette Score evaluates cohesion and separation by computing, for each observation x_i the average intra-cluster distance.

$$a(i) = \frac{1}{|C_{c(i)}| - 1} \sum_{x_j \in C_{c(i)}, j \neq i} ||x_i - x_j|| \tag{3}$$

and the lowest average inter-cluster distance

$$b(i) = \min_{m \neq c(i)} \frac{1}{|C_m|} \sum_{x_j \in C_m} ||x_i - x_j|| \tag{4}$$

then forming the silhouette coefficient

$$s(x_i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \tag{5}$$

with the overall score $\bar{s} = \frac{1}{n} \sum_{i=1}^{n} s(i)$. By comparing the inertia curve and mean silhouette scores for k from 2 to 6, we identified $k = 2$ as providing the best trade-off between compact, well-separated clusters.

3.4. Validation and Visualization

In order to validate the coherence and separation of the two clusters, we employed both quantitative and visual techniques. A silhouette plot was generated by computing the silhouette coefficient $s(x_i)$ for each observation, defined as the ratio of its distance to the nearest non-own cluster over its distance to its own cluster, and then plotting these values in ascending order within each cluster. High average silhouette values indicate that members are appropriately grouped together, while low or negative values reveal ambiguous assignments. Complementing this, we performed dimensionality reduction via Principal Component Analysis (PCA) and t-distributed Stochastic Neighbor Embedding (t-SNE) to project the high-dimensional Q2 and Q3 feature space into two dimensions. By plotting each respondent in PCA and t-SNE space and coloring points by cluster label, we visually confirmed that the two clusters form distinct, well-separated regions.

For centroid interpretation, we translated the cluster centers back into the original Q2 and Q3 variables and rendered these summaries in several chart formats. A heatmap of centroid values highlights, via color intensity, the average binary selection rate for each information source and perception item across clusters. Stacked bar charts then illustrate the proportional composition of sources (Q2) or perceptions (Q3) for each cluster, making it easy to compare relative emphasis on, for example, Internet versus Books/Papers or Problem-Solving versus Job-Replacement concerns. Finally, a radar (spider) chart of the Q3 centroids provides an intuitive profile of each cluster's impact perceptions by plotting the four perception dimensions radially; this format draws immediate attention to which perceptions dominate in each group, such as a strong Problem-Solving orientation versus heightened Dehumanization worries. Together, these validation and visualization steps give both statistical assurance and intuitive insight into the nature of the two respondent segments.

4. Results and Discussion

4.1. Result

Figure 2 plots total within-cluster inertia against the number of clusters k. We observe a sharp decline in inertia from $k = 1$ to $k = 2$, after which the rate of decrease diminishes considerably. This identifiable elbow at $k = 2$ indicates that splitting into two clusters captures the major structure in the data without overfitting; additional clusters yield only marginal gains in compactness, validating our selection of $k = 2$.

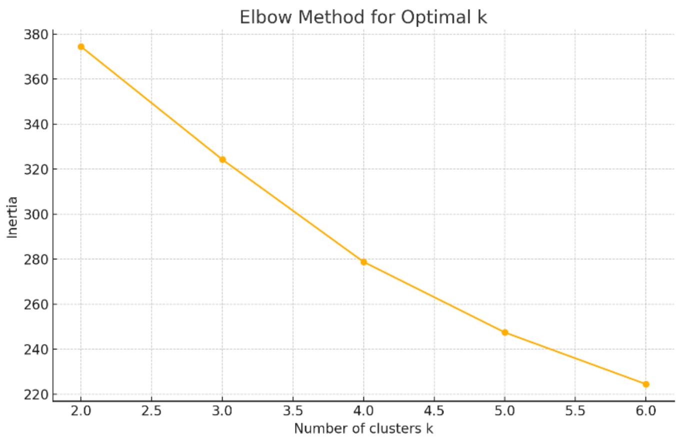

Figure 2. Elbow Method for Optimal k

Figure 3 presents each observation's silhouette coefficient $s(i)$, which measures how well it fits within its assigned cluster versus the nearest alternative. In both Cluster 0 and Cluster 1, most $s(i)$ values are positive, many exceeding 0.2 and reaching up to around 0.5 for Cluster 0 indicating strong cohesion (high intra-cluster similarity) and good separation (low inter-cluster similarity). The absence of substantial negative values confirms few misassignments, further supporting that $k = 2$ produces well-defined clusters.

Figure 3. Silhouette Plot for 2 Clusters

Figure 4 visualizes the results of PCA on the Q2 and Q3 feature set, projecting respondents onto the first two principal components. Points colored by cluster label (purple = Cluster 0, yellow = Cluster 1) form two distinct clouds, underscoring the separation achieved by K-Means. Overlaid loading vectors for each original variable reveal the drivers of this separation: the Problem_solving arrow points toward Cluster 0's region, while Job_replacement and AI_rulling_society arrows align with Cluster 1's domain. This confirms that these perceptions and source variables are the primary dimensions distinguishing the two respondent segments.

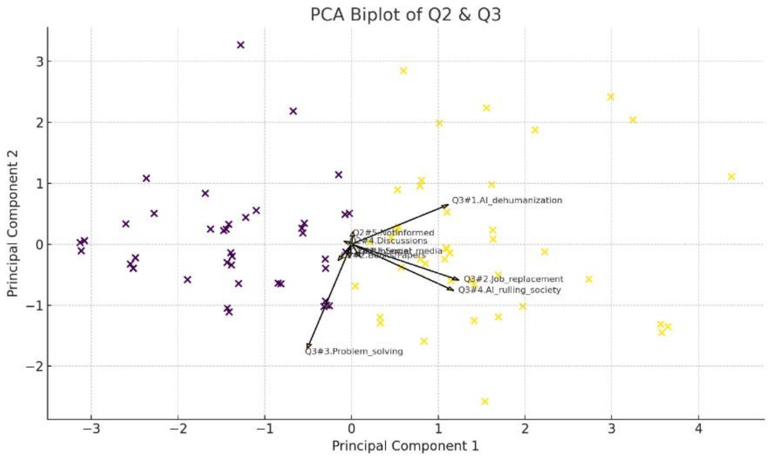

Figure 4. PCA Biplot of Q2 & Q3

Figure 5 displays the average selection rates for each Q2 (information source) and Q3 (impact perception) variable across the two clusters. Cluster 0 shows a particularly high mean on Q3#3.Problem_solving (4.6) with comparatively lower averages on Q3#2.Job_replacement (2.3) and Q3#1.AI_dehumanization (1.9). In contrast, Cluster 1 averages much higher on the risk-oriented perceptions, Job_replacement (4.1) and AI_dehumanization (3.2) and slightly lower on Problem_solving (3.8). The source variables (Q2) also differ: Cluster 0's centroid has higher values for Books/Papers (0.45) and Discussions (0.27), whereas Cluster 1 relies more on Social_media (0.46).

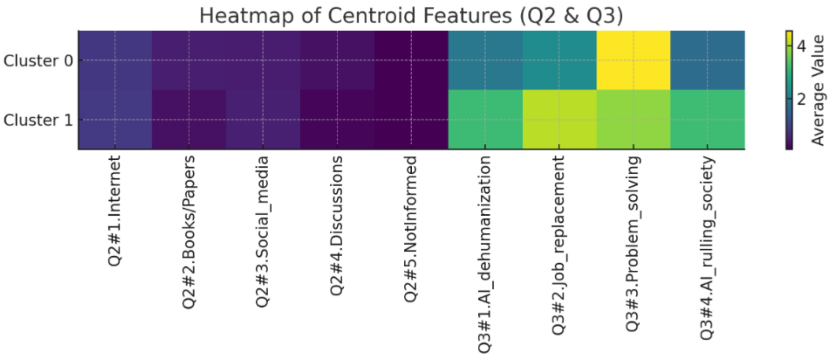

Figure 5. Heatmap of Centroid Features (Q2 & Q3)

Figure 6 presents these Q2 centroids in a bar chart. Both clusters heavily favor the Internet (Cluster 0: 0.80, Cluster 1: 0.83), but Cluster 0 supplements its information with printed or formal channels Books/Papers (0.45) and Discussions (0.27), more than Cluster 1 (Books/Papers 0.26, Discussions: 0.13). Conversely, Cluster 1 shows a noticeably higher reliance on Social_media (0.46 vs. 0.42). This divergence suggests that Optimistic Problem Solvers seek more formal, structured knowledge, while Critical Watchers lean on rapid, user-generated content.

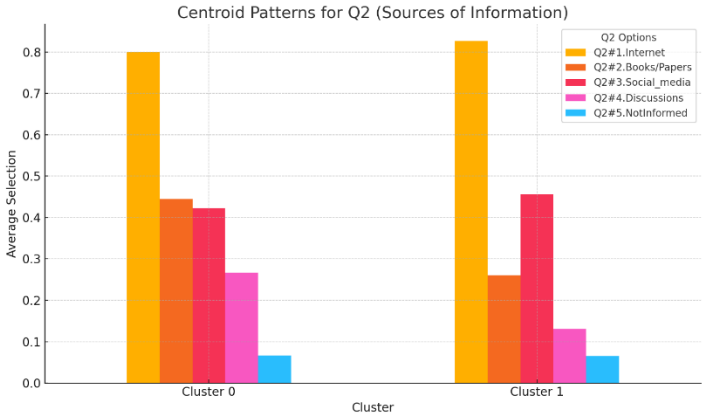

Figure 6. Centroid Patterns for Q2 (Sources of Information)

In Figure 7, the radar (spider) plot maps Q3 centroids on a four-axis chart. Cluster 0's polygon peaks strongly on Problem_solving, creating a pronounced spike, while its values on the risk dimensions remain moderate. Cluster 1 forms a more rounded shape with elevated scores on Job_replacement, AI_dehumanization, and AI_rulling_society, indicating a consistently cautious mindset across multiple concern areas. The contrast in shapes visually encapsulates each cluster's defining perceptual profile.

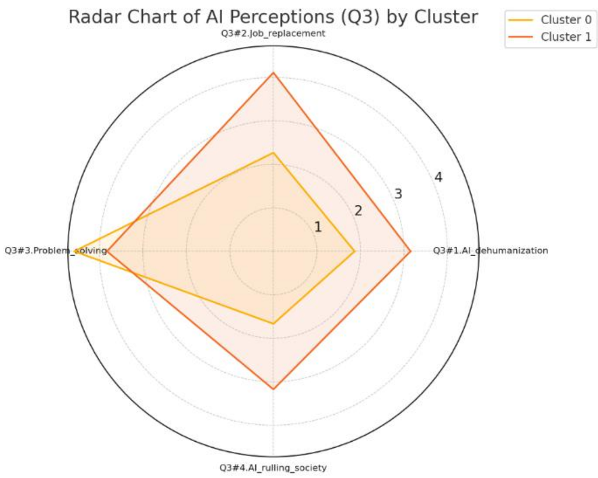

Figure 7. Radar Chart of AI Perceptions (Q3) by Cluster

ISSN 2776-8007

Figure 8 overlays the Q2 centroids in stacked bars, making it easy to compare the relative composition within each cluster. The height of the Internet segment dominates both bars, but the secondary segments differ: Cluster 0's stack includes larger blocks for Books/Papers and Discussions, whereas Cluster 1's stack enlarges the Social_media block. This format underscores how the clusters share common primary channels yet diverge in their secondary information preferences.

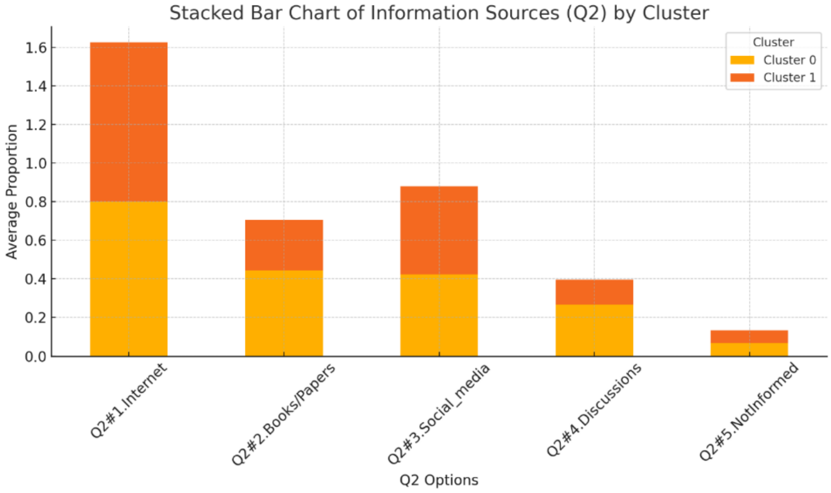

Figure 8. Stacked Bar Chart of Information Sources (Q2) by Cluster

Figure 9 shows boxplots of cumulative GPA for each cluster. Both Cluster 0 and Cluster 1 share a similar median GPA (around 7.7–7.8), indicating comparable central academic performance. However, Cluster 1 exhibits a wider interquartile range and longer whiskers, reflecting greater variability: some Cluster 1 members achieve GPAs above 9.0, while others fall closer to 5.0. In contrast, Cluster 0's GPA scores are more tightly packed between approximately 7.2 and 8.7, suggesting a more academically homogeneous group.

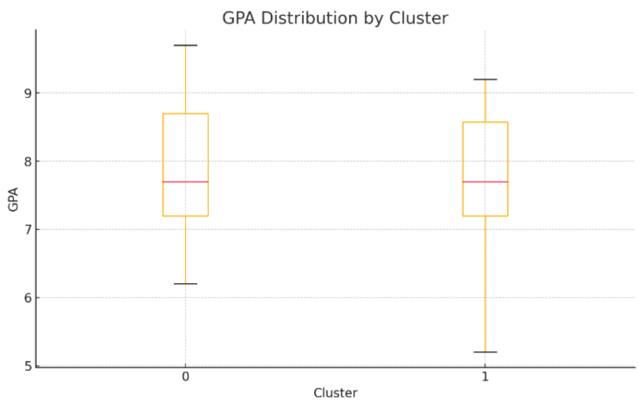

Figure 9. GPA Distribution by Cluster

In figure 10, we present the proportion of male and female respondents within each cluster. Cluster 0 is comprised of roughly 71% male and 29% female students, indicating a strong male predominance. Cluster 1, by comparison, is more gender-balanced with about 59% male and 41% female. This shift suggests that female respondents are relatively more represented among the Critical Watchers segment.

Figure 10. Gender Distribution by Cluster

Figure 11 compares Freshman (Year 1) versus Sophomore (Year 2) proportions across clusters. Cluster 0 has a higher concentration of Sophomores (69%) compared to Freshmen (31%), implying that more advanced students tend to fall into the Optimistic Problem Solvers segment. Cluster 1 is more evenly split 56% Sophomore and 44% Freshman—indicating that both early and later-stage undergraduates share the cautious Critical Watchers profile.

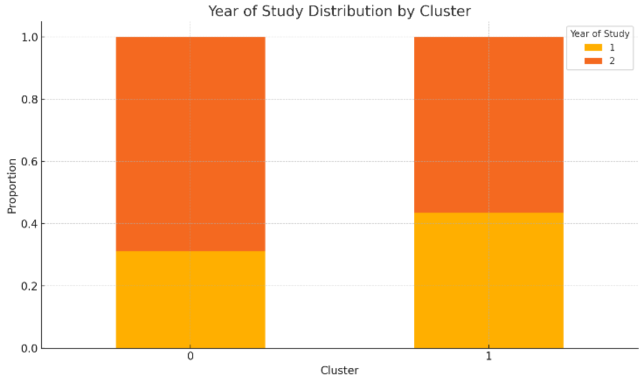

Figure 11. Year of Study Distribution by Cluster

Finally, figure 12 displays the distribution of academic majors in each cluster. Cluster 0 is dominated by students from Major 2 (47%), with fewer from Majors 1 (26%) and 3 (27%). In contrast, Cluster 1 shows a more evenly distributed composition: Major 1 (39%), Major 2 (37%), and Major 3 (24%). This suggests that the Optimistic Problem Solvers segment is more concentrated within a specific discipline (Major 2), whereas the Critical Watchers group spans a broader array of fields.

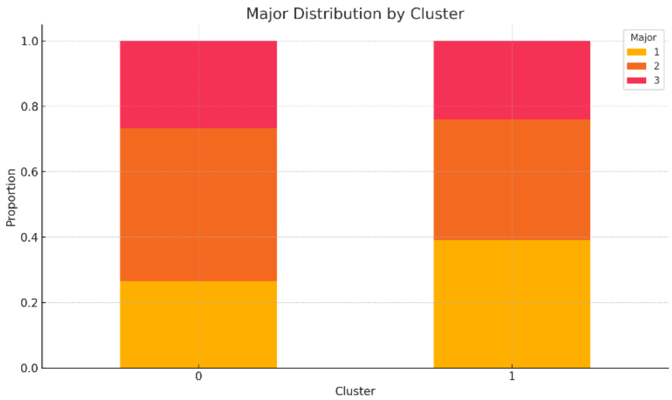

Figure 12. Major Distribution by Cluster.

4.2. Discussion

The Optimistic Problem Solvers cluster, comprising approximately half of the respondents, consistently relies on formal, structured information channels: about 80% consult academic Internet sources, nearly 45% reference books or peer-reviewed papers, and roughly 27% participate in guided discussions. Their impact perception profile peaks strongly on AI's problem-solving potential (average 4.6), while concerns about job replacement (2.3) and dehumanization (1.9) remain modest. Demographically, this cluster skews toward sophomores (69% vs. 31% freshmen), exhibits a tight GPA distribution centered around 7.7–8.0, and is predominantly male (71%) and concentrated in Major 2 (47%). These patterns suggest that more experienced, consistently performing students particularly in quantitatively rigorous discipline's view AI primarily as a tool for enhancing analytical reasoning and complex problem-solving, reflecting findings that high competency learners benefit from data-driven pedagogies [15].

In contrast, the Critical Watchers cluster depends more on informal, rapidly updated sources: 46% use social media compared to 42% in the other cluster, while engagement with books/papers (26%) and discussions (13%) is lower. Their perceptual profile is risk-oriented, with high averages for job displacement (4.1) and dehumanization (3.2), and a moderate score for problem-solving (3.8). This cluster is more gender-balanced (59% male, 41% female), includes a higher proportion of freshmen (44% vs. 56% sophomores), and shows greater GPA variability (5.2–9.0). Their evenly distributed majors indicate that students from less computational fields harbor greater skepticism toward AI, perhaps due to limited curricular emphasis on algorithmic literacy. This aligns with network-based segmentation research showing that groups defined by informal, peer-driven interactions exhibit greater skepticism toward authoritative narratives [14].

The stark divergence between clusters underscores the inadequacy of a one size fits all approach to AI education. For Optimistic Problem Solvers, advanced, research-oriented modules can deepen engagement: hands-on workshops in machine-learning model development, data visualization, and algorithmic bias mitigation align with their problem-solving orientation. These students mirror the technology compassionates identified in hospitality studies, who engage deeply when content resonates with their analytical predilections [19].

Conversely, Critical Watchers require trust-building outreach delivered through the channels they frequent social media, AI chatbots, and peer forums. Bite-sized, narrative-driven content explaining how AI models arrive at decisions can directly address their opacity concerns. Interactive Q&A sessions with AI developers and ethics-in-AI seminars can humanize technology, counteract misinformation, and provide safe spaces for voicing concerns, gradually shifting perceptions toward informed acceptance.

Moreover, proactively integrating ethical discussions on topics like job displacement and dehumanization into introductory AI literacy courses can legitimize Critical Watchers reservations and model responsible AI use. Embedding reflective assignments that ask students to analyze real-world AI failures may further engage risk-oriented learners by validating their concerns and demonstrating mitigations.

Our segmentation both converges with and extends existing literature. In vocational education, Faisal et al. [21] applied K-Means to group students by ICT competency, enabling targeted instruction that improved engagement paralleling our Optimistic Problem Solvers. Qiao [22] showed that cluster-informed civic education interventions increase student participation, supporting our recommendation for tailored outreach to Critical Watchers. Li et al. [24] reported that clinicians with advanced training trust AI more, further reinforcing the link between formal knowledge acquisition and acceptance. By integrating both information-source behaviors and perception dimensions, our study provides a unified segmentation framework that bridges these domain-specific insights and applies them to a higher-education setting.

Several limitations temper our conclusions. The modest sample size (n = 91) from a single institution constrains external validity; replication with larger, multi-institutional cohorts is needed to test the stability of these clusters across diverse academic environments. Our binary encoding of multi-response items simplifies nuanced behaviors future research should explore ordinal or weighted encoding to capture intensity of preferences. The cross-sectional design precludes analysis of how cluster memberships evolve over time, Qaladi et al. [12] emphasize the importance of longitudinal tracking to understand attitude trajectories. Incorporating mixed methods such as in-depth interviews could enrich quantitative clusters with deeper insights into students' motivations, emotional drivers, and contextual factors.

Future work should implement longitudinal surveys to monitor movement between Optimistic and Critical segments following specific interventions, thereby empirically testing the efficacy of tailored curricula. Integrating learning analytics clickstream data from AI enhanced platforms could refine cluster definitions and enable real-time personalization. Additionally, comparative studies across cultural contexts would illuminate how socio-cultural factors influence information behaviors and AI perceptions.

This research advances educational data mining by demonstrating that dual-dimensional segmentation combining source behaviors and impact perceptions yields actionable profiles for designing differentiated AI curricula, communication campaigns, and governance protocols. Practically, institutions can deploy modular AI literacy programs, a deep-dive track for Optimistic Problem Solvers and a trust-building track for Critical Watchers, delivered via appropriate channels. Ongoing cluster monitoring can guide iterative content refinement, ensuring responsiveness to evolving student needs.

In sum, segmenting AI survey respondents into two distinct clusters provides a robust, data-driven foundation for personalized AI education strategies, highlighting the imperative of aligning pedagogical design with students' information habits, perceptual orientations, and demographic contexts to foster equitable and effective AI adoption in higher education.

5. Conclusion

This study applied K-Means clustering to segment 91 university students based on their information-seeking behaviors (Q2) and impact perceptions (Q3) of artificial intelligence (AI). Two distinct clusters Optimistic Problem Solvers and Critical Watchers emerged, each characterized by unique combinations of preferred information sources, perceptual emphases, and demographic profiles. The Optimistic Problem Solvers heavily rely on formal, structured channels (e.g., scholarly Internet resources, peer-reviewed papers, guided discussions) and express strong confidence in AI's problem-solving capabilities while harboring relatively low concerns about job displacement or dehumanization. In contrast, the Critical Watchers depend more on informal, rapidly updated media (e.g., social media, general web searches) and display heightened apprehension regarding AI's socio-economic and ethical risks, alongside more varied academic standings and broader disciplinary representation. These findings validate the dual-dimensional segmentation framework combining source behavior with perceptual orientation as an effective approach for uncovering meaningful user profiles that extend beyond traditional demographic or competency-based groupings. The clear divergence between clusters underscores the limitations of one-size-fits-all AI education and highlights the need for differentiated instructional strategies: deep-dive, research-oriented modules for the Optimistic Problem Solvers and trust-building, narrative-driven outreach via familiar channels for the "Critical Watchers." Such tailored interventions promise to enhance engagement, build confidence, and address ethical concerns in a manner aligned with each group's preferences.

Despite its contributions, this research is limited by its single-institution sample, binary encoding of multi-response items, and cross-sectional design. Future studies should employ larger, more diverse cohorts, incorporate graded preference measures, and adopt longitudinal and mixed-methods approaches to track profile evolution and assess intervention efficacy. Integrating real-time learning analytics may further refine segmentation and enable dynamic personalization of AI literacy programs. By delineating the contours of AI information ecosystems and perceptual landscapes, this study offers both methodological advances in educational data mining and practical guidance for higher education stakeholders. Implementing modular AI literacy tracks one emphasizing technical depth and another fostering transparent, ethics-oriented engagement can ensure that AI education is equitable, effective, and responsive to the diverse needs of today's learners.

6. Declarations

6.1. Author Contributions

Conceptualization: E., S.P.S., M.T.; Methodology: E., M.T.; Software: E.; Validation: S.P.S., M.T.; Formal Analysis: E.; Investigation: E.; Resources: S.P.S., M.T.; Data Curation: E.; Writing – Original Draft Preparation: E.; Writing – Review and Editing: S.P.S., M.T.; Visualization: E.; All authors have read and agreed to the published version of the manuscript.

6.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

ISSN 2776-8007

6.3. Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

6.4. Institutional Review Board Statement

Not applicable.

6.5. Informed Consent Statement

Not applicable.

6.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] X. Chen, H. Xie, S. J. Qin, F. L. Wang, and Y. Hou, "Artificial Intelligence-Supported Student Engagement Research: Text Mining and Systematic Analysis," *Eur. J. Educ.*, vol. 60, no. 1, pp. 1–16, 2025, doi: 10.1111/ejed.70008.

[2] C. K. Y. Chan and W. Hu, "Students' Voices on Generative AI: Perceptions, Benefits, and Challenges in Higher Education," *Int. J. Educ. Technol. High. Educ.*, vol. 20, no. 1, pp. 1–18, 2023, doi: 10.1186/s41239-023-00411-8.

[3] G. A. L. Quinde, M. Y. T. Muñoz, J. M. R. Suárez, R. E. P. Villarreal, W. A. Z. Vélez, and A. A. D. P. Laínez, "Perception of University Students on the Use of Artificial Intelligence (AI) Tools for the Development of Autonomous Learning," *Rev. Gestão Soc. E Ambient.*, vol. 18, no. 2, pp. 1–18, 2024, doi: 10.24857/rgsa.v18n2-136.

[4] H. Jo and Y. Bang, "Analyzing ChatGPT Adoption Drivers With the TOEK Framework," *Sci. Rep.*, vol. 13, no. 1, pp. 1–17, 2023, doi: 10.1038/s41598-023-49710-0.

[5] A. Q. Sarwari and H. M. Adnan, "The Effectiveness of Artificial Intelligence (AI) on Daily Educational Activities of Undergraduates in a Modern and Diversified University Environment," *Adv. Mob. Learn. Educ. Res.*, vol. 4, no. 1, pp. 927–930, 2024, doi: 10.25082/amler.2024.01.004.

[6] L. Sun and L. Zhou, "Does Generative Artificial Intelligence Improve the Academic Achievement of College Students? A Meta-Analysis," *J. Educ. Comput. Res.*, vol. 62, no. 7, pp. 1676–1713, 2024, doi: 10.1177/07356331241277937.

[7] M. Mustopa, N. Nasikhin, R. Chamami, H. Nihayah, M. R. Habibullah, and A. Manshur, "Challenges in Artificial Intelligence Development in Higher Education in China, India, and Indonesia: International Students' Perspectives," *Int. J. Learn. Teach. Educ. Res.*, vol. 23, no. 2, pp. 354–373, 2024, doi: 10.26803/ijlter.23.2.17.

[8] R. Kaur, D. Gupta, M. Madhukar, A. Singh, M. Abdelhaq, R. Alsaqour, J. Brenosa, and N. Goyal, "E-Learning Environment Based Intelligent Profiling System for Enhancing User Adaptation," *Electronics*, vol. 11, no. 20, pp. 3354, 2022, doi: 10.3390/electronics11203354.

[9] M. Kordahi, "Ontology for the User-Learner Profile Personalizes the Search Analysis of Online Learning Resources," *Inf. Technol. Libr.*, vol. 41, no. 2, pp. 1–22, 2022, doi: 10.6017/ital.v41i2.13601.

[10] T. Blanco, S. Martín-Segura, J. L. d. Larrinzar, R. Béjar, and F. J. Zarazaga-Soria, "First Steps Toward Voice User Interfaces for Web-Based Navigation of Geographic Information: A Spanish Terms Study," *Appl. Sci.*, vol. 13, no. 4, pp. 2083, 2023, doi: 10.3390/app13042083.

[11] M. A. Kraft and A. Bolves, "Can Technology Transform Communication Between Schools, Teachers, and Parents? Evidence From a Randomized Field Trial," *Educ. Financ. Policy*, vol. 17, no. 3, pp. 479–510, 2022, doi: 10.1162/edfp_a_00344.

[12] O. Qaladi, M. Alshammari, and A. Almalki, "Artificial Intelligence (AI) in Nursing Administration: Challenges and Opportunities," *PLoS One*, vol. 20, no. 4, pp. e0319588, 2025, doi: 10.1371/journal.pone.0319588.

[13] A. S. Valerio, "Anticipating the Impact of Artificial Intelligence in Higher Education: Student Awareness and Ethical Concerns in Zamboanga City, Philippines," *Cogniz. J. Multidiscip. Stud.*, vol. 4, no. 6, pp. 408–418, 2024, doi: 10.47760/cognizance.2024.v04i06.024.

[14] L. Ma and Y. Zhang, "Three Social-Mediated Publics in Digital Activism: A Network Perspective of Social Media Public Segmentation," *Soc. Media + Soc.*, vol. 8, no. 2, pp. 1–10, 2022, doi: 10.1177/20563051221094775.

[15] W.M.Zhou, "Analysis of Student Behavior in English Online Education Under the Perspective of Big Data," *Int. J. Decis. Support Syst. Technol.*, vol. 17, no. 1, pp. 1–22, 2025, doi: 10.4018/ijdsst.378427.

[16] M. Kinnunen, H. Homi, and A. Honkanen, "Live Music Consumption of the Adolescents of Generation Z," *Etnomusikol. Vuosik.*, vol. 34, no. December, pp. 65–92, 2022, doi: 10.23985/evk.114885.

[17] M. Mukoka, T. C. Msosa, H. H. Twabi, R. Semphere, M. Nliwasa, G. Harling, A. Price, K. Fielding, and A. T. Choko, "The Key Design Features and Effectiveness of Social Network Interventions for HIV Testing and Linkage Services in Low- and Middle-income Countries: A Systematic Review and Meta-analysis," *J. Int. Aids Soc.*, vol. 28, no. 5, pp. e26458, 2025, doi: 10.1002/jia2.26458.

[18] A. A. Safeer, "Harnessing the Power of Brand Social Media Marketing on Consumer Online Impulse Buying Intentions: A Stimulus-Organism-Response Framework," *J. Prod. Brand Manag.*, vol. 33, no. 5, pp. 533–544, 2024, doi: 10.1108/jpbm-07-2023-4619.

[19] T. Demirçiftçi, A. Bilgihan, M. Erdem, and Ş. Baloğlu, "Examining the Role of Personality Traits in Guestroom Technologies," *J. Hosp. Tour. Insights*, vol. 7, no. 1, pp. 352–370, 2023, doi: 10.1108/jhti-11-2022-0529.

[20] A. J. E. Sakalessy and H. D. Purnomo, "Assessing Employee Performance in the Information Technology Department Using K-Means Clustering: A Case Study Approach," *J. Inf. Syst. Informatics*, vol. 6, no. 1, pp. 170–186, 2024, doi: 10.51519/journalisi.v6i1.653.

[21] M. Faisal, N. Nurdin, F. Fajriana, and Z. Fitri, "Information and Communication Technology Competencies Clustering for Students for Vocational High School Students Using K-Means Clustering Algorithm," *Int. J. Eng. Sci. Inf. Technol.*, vol. 2, no. 3, pp. 111–120, 2022, doi: 10.52088/ijesty.v2i3.318.

[22] Y. Qiao, "The Combination Application of Civic Education and Big Data Internet Technology Based on the Background of College Education Reform," *Appl. Math. Nonlinear Sci.*, vol. 9, no. 1, pp. 1–17, 2023, doi: 10.2478/amns.2023.2.00168.

[23] F. Ibrahim, J. Münscher, M. Daseking, and N.-T. Telle, "The Technology Acceptance Model and Adopter Type Analysis in the Context of Artificial Intelligence," *Front. Artif. Intell.*, vol. 7, no. January, pp. 1–14, 2025, doi: 10.3389/frai.2024.1496518.

[24] N. Li, L. Li, M. Xie, and B. Yalvaç, "Develop and Disseminate Hands-on Lab Materials of Privacy Concepts and Technologies to Educators," *J. Colloq. Inf. Syst. Secur. Educ.*, vol. 11, no. 1, pp. 1–9, 2024, doi: 10.53735/cisse.v11i1.182.

[25] S. Edgington and K. Kasztelnik, "The Ethical Considerations of Business Artificial Intelligence Exploration Through the Lenses of the Global AI Technology Acceptance Model," *J. Strateg. Innov. Sustain.*, vol. 19, no. 1, pp. 13–22, 2024, doi: 10.33423/jsis.v19i1.6749.

[26] A. Schepart, A. Burton, L. Durkin, A. Fuller, E. Charap, R. Bhambri, and F. S. Ahmad, "Artificial Intelligence–enabled Tools in Cardiovascular Medicine: A Survey of Current Use, Perceptions, and Challenges," *Cardiovasc. Digit. Heal. J.*, vol. 4, no. 3, pp. 101–110, 2023, doi: 10.1016/j.cvdhj.2023.04.003.

[27] P. Jiang, W. Niu, Q. Wang, R. Yuan, and K. Chen, "Understanding Users' Acceptance of Artificial Intelligence Applications: A Literature Review," *Behav. Sci. (Basel).*, vol. 14, no. 8, pp. 1–30, 2024, doi: 10.3390/bs14080671.

[28] L. Shinners, C. Aggar, A. S. Stephens, and S. Grace, "Healthcare Professionals' Experiences and Perceptions of Artificial Intelligence in Regional and Rural Health Districts in Australia," *Aust. J. Rural Health*, vol. 31, no. 6, pp. 1203–1213, 2023, doi: 10.1111/ajr.13045.

[29] X. Liu and Z. Shi, "Served or Exploited: The Impact of Data Capture Strategy on Users' Intention to Use Selected AI Systems," *J. Consum. Behav.*, vol. 24, no. 2, pp. 1003–1016, 2025, doi: 10.1002/cb.2428.

[30] A. J. Adetayo, "Artificial Intelligence Chatbots in Academic Libraries: The Rise of ChatGPT," *Libr. Hi Tech News*, vol. 40, no. 3, pp. 18–21, 2023, doi: 10.1108/lhtn-01-2023-0007.

[31] R. Yu, S. Lee, J. Xie, S. M. Billah, and J. M. Carroll, "Human–AI Collaboration for Remote Sighted Assistance: Perspectives From the LLM Era," *Futur. Internet*, vol. 16, no. 7, pp. 254, 2024, doi: 10.3390/fi16070254.