A Clustering-Based Framework for AI User Profiling in Intelligent Systems

Mohammed O. Alannsary
The institute of Public Administration (Saudi Arabia), annsary@ipa.edu.sa
https://orcid.org/0000-0003-2254-1441

Received: 22 /04 /2026 Accepted: 15/05 /2026 Published: 01 /06 /2026

Abstract:

Artificial intelligence (AI) systems increasingly depend on user modeling to provide adaptive functionality and personalized interactions. Although prior studies have examined AI adoption using descriptive analysis or supervised prediction models, comparatively limited attention has been given to robust unsupervised profiling of AI users across multiple usage contexts. This study proposes a clustering-based framework for profiling AI users in professional environments using a multidimensional dataset derived from structured survey responses. The framework combines data preprocessing, unsupervised learning, dimensionality reduction, and cluster validation to generate interpretable user profiles from behavioral and contextual variables. Using a dataset of 62 participants, the analysis shows that a two-cluster solution provides the most suitable segmentation of the data. The identified clusters represent two distinct user profiles: low-integration users and high-integration users. The results indicate that the primary separation between users is driven by task-oriented AI usage in professional and practical contexts rather than by general familiarity with AI tools. The study further shows that structured survey data can be transformed into computational user models suitable for intelligent systems design. By linking user segmentation to personalization needs, the proposed framework contributes to data mining, user modeling, and adaptive AI system development.

Keywords: Artificial Intelligence; User Profiling; Clustering; Unsupervised Learning; Intelligent Systems; Personalization.

INTRODUCTION

The rapid advancement of artificial intelligence (AI) technologies has significantly reshaped the design and functionality of modern digital systems. AI is increasingly embedded in professional environments, supporting tasks such as content generation, decision-making, and knowledge retrieval, particularly in regions undergoing rapid digital transformation such as Saudi Arabia (SDAIA, 2024). As these systems become more pervasive, there is a growing need to design them in ways that accommodate diverse user behaviors and usage patterns. In this context, user modeling has emerged as a fundamental component of intelligent systems, enabling adaptive functionality and personalized interactions (Brusilovsky, 2001; Fischer, 2001; Kobsa, 2007; Gauch et al., 2007).

Traditional research on AI adoption has primarily focused on descriptive analysis or predictive modeling approaches that aim to explain or forecast user behavior. While such approaches provide valuable insights, they often rely on predefined categories or assumptions regarding user groups. This can limit their ability to capture the inherent complexity and heterogeneity of user behavior, particularly in environments where AI usage spans multiple contexts, such as personal and professional domains. As a result, there is a need for computational approaches that can uncover latent structures in user data without imposing prior assumptions (Jain, 2010; Kergroach, 2025; OECD, 2025; Capel & Brereton, 2023).

Unsupervised learning techniques, particularly clustering, offer a powerful alternative for analyzing user behavior. By grouping users based on similarity in their feature representations, clustering enables the discovery of hidden patterns and natural groupings within data. These techniques have been widely applied in areas such as recommendation systems, human–computer interaction, and adaptive learning environments, where identifying user segments can significantly enhance system performance and usability. However, the application of clustering methods to AI usage behavior—especially using structured survey-based datasets—remains relatively underexplored (Wedel & Kamakura, 2020; Ricci et al., 2022; Everitt et al., 2021; JOLLIFFE & CADIMA, 2016).

AI usage behavior is inherently multidimensional, involving not only frequency of use but also the types of tasks performed, the contexts in which AI is applied, and the barriers that influence adoption. Users may interact with AI differently in personal settings compared to professional environments, and factors such as perceived reliability, privacy concerns, and technical skills can further shape these interactions. Capturing these dimensions requires a comprehensive representation of user behavior that goes beyond simple usage metrics (Amershi et al., 2019; Raisch & Krakowski, 2021; Tarafdar et al., 2022; Abdi & Williams, 2010).

This study addresses these challenges by proposing a clustering-based framework for profiling AI users in professional environments. The framework treats user data as a multidimensional feature space and applies unsupervised learning techniques to identify distinct behavioral profiles. Unlike traditional approaches that focus on prediction, this study emphasizes the discovery and interpretation of latent user groups, providing a data-driven basis for understanding AI usage patterns. This conceptualization is

Volume 7, Issue 26 (2026) p428- 449

operationalized through a structured multi-phase framework that integrates data preparation, clustering, validation, and system-level application, as illustrated in (Fig. 1).

The analysis is conducted using a real-world dataset consisting of responses from 62 participants, capturing personal and professional AI usage, perceived barriers, and contextual factors (Alannsary, 2025). The results reveal that user behavior is best represented by a two-cluster structure, distinguishing between low-integration users and high-integration users. This finding suggests that AI adoption is not a continuous process but rather a structured phenomenon characterized by distinct usage patterns.

The contributions of this study are threefold. First, this work proposes a hybrid clustering framework that integrates feature engineering, K-means clustering, hierarchical validation, and dimensionality reduction to systematically profile AI users based on behavioral patterns.

Second, the study provides an empirical analysis of AI usage behavior using real-world survey data, identifying distinct user groups characterized by varying levels of personal and professional AI integration as well as perceived barriers.

Third, and most importantly, this research offers a practical contribution by demonstrating how data-driven user profiling can be translated into actionable system-level personalization strategies, enabling adaptive interfaces, targeted recommendations, and context-aware support in AI-driven systems.

These contributions collectively bridge the gap between analytical user profiling and the design of adaptive intelligent systems.

By bridging the gap between user behavior analysis and intelligent system design, this study contributes to the fields of data mining, user modeling, and AI-driven personalization, offering a scalable approach for profiling users in emerging AI-enabled environments.

Related Work

User modeling and personalization have become central research themes in modern intelligent systems, particularly in applications involving human–AI interaction. As AI systems increasingly operate in dynamic and user-centric environments, the ability to construct accurate representations of users has become essential for enabling adaptive behavior and improving system effectiveness. Recent studies emphasize the role of data-driven user modeling, where behavioral and contextual data are used to infer user characteristics and support personalized system responses (Brusilovsky, 2001; Fischer, 2001; Kobsa, 2007; Gauch et al., 2007).

Unsupervised learning techniques, particularly clustering, have been widely adopted for user profiling due to their ability to identify latent structures in data without requiring labeled inputs. Clustering-based approaches have been successfully applied in domains such as recommendation systems, adaptive learning platforms, and human–computer interaction. By grouping users based on similarity in their interaction patterns or preferences, these methods enable the identification of user segments that can be leveraged for personalization. Recent work has shown that clustering can be used to construct user personas that improve system usability and support adaptive interface design. Recent clustering-based studies have shown that personas and user segments can be derived from behavioral data to support adaptive system design across educational, civic, and interaction-centered contexts (VENKATESH et al., 2003; Amershi et al., 2019; Wedel & Kamakura, 2020; Abdi & Williams, 2010).

In parallel, advances in data mining have highlighted the importance of multidimensional feature representation in improving clustering performance. Rather than relying on a single type of data, modern approaches integrate behavioral, contextual, and demographic features to capture the complexity of user behavior. This integration has been shown to enhance the interpretability and robustness of clustering results, particularly in environments where user behavior is influenced by multiple interacting factors (Han et al., 2011; Everitt et al., 2021; Géron, 2022).

Dimensionality reduction techniques, such as principal component analysis (PCA), are often used in conjunction with clustering to improve both computational efficiency and interpretability. By projecting high-dimensional data into a lower-dimensional space, these methods help identify the dominant factors that drive user differentiation. Recent studies demonstrate that combining clustering with dimensionality reduction can provide deeper insights into behavioral patterns and support more effective user modeling (Han et al., 2011; Abdi & Williams, 2010).

Despite these advancements, much of the existing research relies on system-generated interaction data, such as logs or clickstream data, which capture user behavior in specific system contexts. In contrast, survey-based datasets offer a complementary perspective by capturing not only usage behavior but also user perceptions, attitudes, and barriers. However, the application of clustering techniques to such datasets remains relatively limited, particularly in the context of AI usage across both personal and professional domains (Ricci et al., 2022; Capel & Brereton, 2023; Alannsary, 2025).

Moreover, many studies focus on supervised approaches for predicting technology adoption, often framing the problem as a classification or regression task. While these approaches are useful for forecasting behavior, they do not necessarily reveal the underlying structure of user populations. Unsupervised approaches, by contrast, enable the discovery of natural groupings without predefined labels, making them particularly suitable for exploratory analysis in emerging domains such as AI usage (Dwivedi et al., 2023; Kergroach, 2025; OECD, 2025; Everitt et al., 2021).

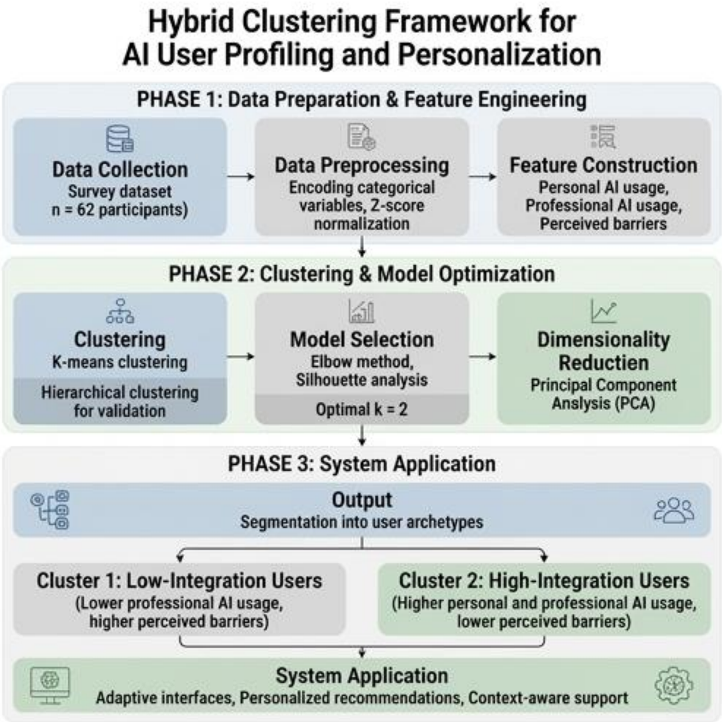
This study builds on existing work in data mining and user modeling by applying clustering techniques to a multidimensional survey dataset that captures both behavioral and perceptual aspects of AI usage. By integrating clustering with validation and dimensionality reduction, the proposed approach extends prior research and demonstrates how structured survey data can be used to construct robust and interpretable user profiles. These profiles provide a foundation for designing adaptive and personalized AI systems that respond to diverse user needs.

Methodology

This study adopts a data-driven approach based on unsupervised machine learning to model AI usage behavior among professionals (Murphy, 2022). The objective is to identify latent user structures within a multidimensional dataset and derive interpretable user profiles that can support adaptive and personalized AI systems. To ensure robustness and methodological rigor, the proposed framework integrates data preprocessing, multiple clustering techniques, dimensionality reduction, and validation procedures (Han et al., 2011; Gauch et al., 2007; Hastie et al., 2021).

To provide a comprehensive overview of the proposed approach, the overall workflow of the hybrid clustering framework is illustrated in (Fig. 1). The framework integrates data preparation, feature engineering, clustering, validation, and system-level application into a unified pipeline for AI user profiling and personalization.

Fig. 1. Overview of the proposed hybrid clustering framework for AI user profiling and personalization, illustrating the stages of data preparation, clustering and model optimization, and system-level application.



As shown in (Fig. 1), the framework is structured into three main phases: data preparation and feature engineering, clustering and model optimization, and system application. This structure enables a systematic transformation of raw survey data into interpretable user profiles that can support adaptive intelligent systems.

Dataset and Feature Representation

The dataset used in this study consists of 62 valid responses collected through a structured questionnaire administered to professionals in a training environment. The dataset captures multiple aspects of AI usage behavior, including personal usage, professional application, and perceived barriers (Capel & Brereton, 2023; Alannsary, 2025).

Each participant is represented as a feature vector in a multidimensional space. The feature set is organized into three main categories:

- Personal AI Usage Features:

  Variables representing the use of AI in everyday activities, such as information retrieval, content generation, and decision support.

- Professional AI Usage Features:

Volume 7, Issue 26 (2026) p428- 449

Variables capturing the use of AI in work-related tasks, including content preparation, evaluation, instructional support, and task automation.

- Barrier Features:

Variables reflecting perceived obstacles to AI adoption, such as lack of technical skills, privacy concerns, trust in AI outputs, and fear of job displacement. Similar barrier dimensions-especially trust, skills, and privacy-have also been reported in recent AI adoption studies (Dwivedi et al., 2023; Kergroach, 2025; Capel & Brereton, 2023).

Demographic attributes are included to support interpretation of clustering results but are not treated as primary drivers of clustering.

Data Preprocessing

Prior to clustering, the dataset undergoes several preprocessing steps to ensure consistency and suitability for machine learning analysis.

First, categorical variables are transformed into numerical representations using appropriate encoding schemes. Ordinal responses are mapped to ordered numerical values, while nominal variables are encoded using one-hot encoding where necessary.

Second, all features are standardized using Z-score normalization, defined as:

$$z = \frac{x - \mu}{\sigma}$$

where x is the original value, μ is the feature mean, and σ is the standard deviation. This step ensures that all variables contribute equally to the clustering process and prevents features with larger scales from dominating distance calculations.

Finally, the dataset is inspected for missing or inconsistent values. Any incomplete entries are removed to maintain data integrity (Han et al., 2011; Hastie et al., 2021).

Clustering Framework

Clustering techniques such as K-means remain widely used for user segmentation tasks (Jain, 2010), particularly within contemporary data-driven systems (Hastie et al., 2021). To identify latent user groups, the study employs a hybrid clustering framework that combines multiple unsupervised learning techniques.

K-means Clustering

K-means clustering is used as the primary segmentation method. The algorithm partitions the dataset into k clusters by minimizing the within-cluster sum of squared distances between data points and their corresponding centroids. Formally, the objective function is defined as:

$$\sum_{i=1}^{k}\sum_{x\in C_i}\|x-\mu_i\|^2$$

Type equation here.

where C_i represents cluster i and μ_i is its centroid.

K-means is selected due to its computational efficiency and suitability for numerical feature spaces (Han et al., 2011; Everitt et al., 2021).

Hierarchical Clustering

To validate the structure identified by K-means, hierarchical clustering is also applied. This method builds a dendrogram representing nested groupings of data points based on similarity measures.

The consistency between K-means and hierarchical clustering results is used as an indicator of cluster stability. Agreement between methods strengthens confidence in the identified segmentation.

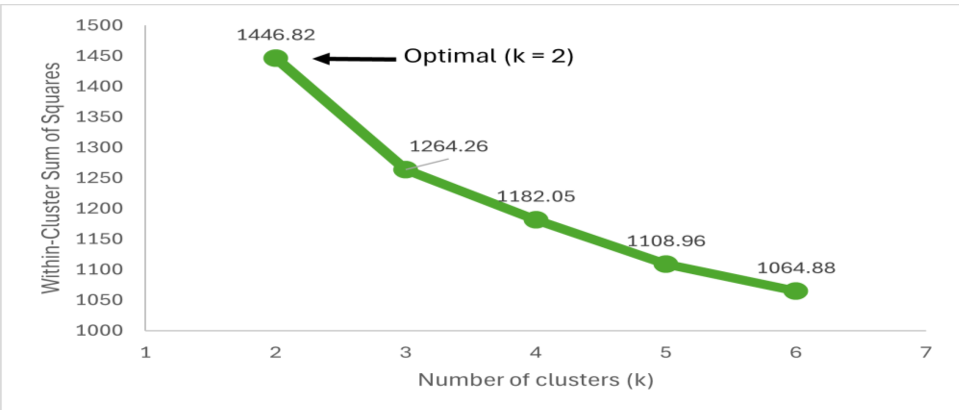
Determination of Optimal Number of Clusters

Selecting an appropriate number of clusters is a critical step in clustering analysis. This study employs two complementary methods:

- Elbow Method:

  The within-cluster sum of squares (WCSS) is computed for different values of k. The optimal number of clusters is identified at the point where the rate of decrease sharply changes. As shown in (Fig. 2), the elbow point is observed at k = 2, indicating the optimal number of clusters.

Fig. 2. Elbow method for determining the optimal number of clusters (k). The within-cluster sum of squares (WCSS) shows a clear inflection point at k = 2, suggesting the optimal clustering configuration.



- Silhouette Analysis:

  The Silhouette method (Rousseeuw, 1987), as implemented in modern machine learning workflows (Hastie et al., 2021), was used to evaluate cluster cohesion and separation. The silhouette score measures the degree of separation between clusters. It is defined as:

$$s = \frac{b - a}{\max(a, b)}$$

where a is the average intra-cluster distance and b is the average nearest-cluster distance. Higher values indicate better clustering quality. (Fig. 3) confirms that k = 2 yields the highest silhouette score, supporting the selection of the optimal clustering configuration.

Fig. 3. Silhouette scores for different numbers of clusters. The highest score is achieved at k = 2, confirming the optimal clustering solution.

Based on these criteria, the two-cluster solution is selected as the most appropriate representation of the dataset.

Dimensionality Reduction

To analyze the structure of the feature space and improve interpretability, principal component analysis (PCA) is applied. PCA (Jolliffe & Cadima, 2016), is a widely adopted dimensionality reduction technique in modern analytics pipelines. PCA transforms the original high-dimensional data into a set of orthogonal components that capture the maximum variance (Han et al., 2011; Abdi & Williams, 2010). This alignment between dominant principal components and clustering structure indicates that the identified user segments are not arbitrary, but are driven by underlying latent dimensions in the data, further strengthening the validity of the clustering results.

This step serves two purposes:

1) Reducing dimensionality for visualization and analysis
2) Identifying dominant factors influencing user behavior

The resulting components provide insight into the underlying structure of the dataset and support interpretation of clustering results.

Cluster Validation and Stability

To ensure the reliability of the clustering results, multiple validation strategies are employed:

- **Silhouette scores** are used to evaluate cluster cohesion and separation
- **Cross-method comparison** between K-means and hierarchical clustering is performed
- **Consistency of cluster assignments** is examined to assess stability

These validation steps ensure that the identified clusters are not artifacts of a single method but represent meaningful structures in the data.

User Profiling Model

Based on the clustering results, a formal user profiling model is defined as:

$$User\ Profile = \{Usage\ Intensity, Professional\ Integration, Barrier\ Level\}$$

This model provides a structured representation of user behavior that can be directly integrated into intelligent systems. By mapping users to specific profiles, systems can adapt functionality and interaction strategies based on user characteristics.

System-Level Implications

The proposed methodology establishes a direct link between data mining outputs and intelligent system design. The derived user profiles can be used to inform:

- Adaptive user interfaces
- Personalized recommendation mechanisms
- Context-aware decision support

By incorporating clustering-based user modeling, AI systems can dynamically adjust behavior to better match user needs, improving both usability and system performance.

Volume 7, Issue 26 (2026) p428- 449

Results and Analysis

This section presents the results of the clustering analysis and evaluates the structure of the identified user groups. The findings are organized into three parts: determination of the optimal number of clusters, cluster distribution, and interpretation of cluster characteristics.

Determination of Optimal Number of Clusters

To identify the optimal number of clusters, K-means clustering was performed for values of k ranging from 2 to 6. The evaluation was conducted using both the elbow method and silhouette analysis.

The within-cluster sum of squares (WCSS) decreased monotonically as the number of clusters increased, with values of 1446.82 for $k = 2$, 1264.26 for $k = 3$, 1182.05 for $k = 4$, 1108.96 for $k = 5$, and 1064.88 for $k = 6$. Although this reduction indicates improved compactness with higher values of k, the rate of decrease becomes progressively smaller, suggesting diminishing returns beyond for $k = 2$.

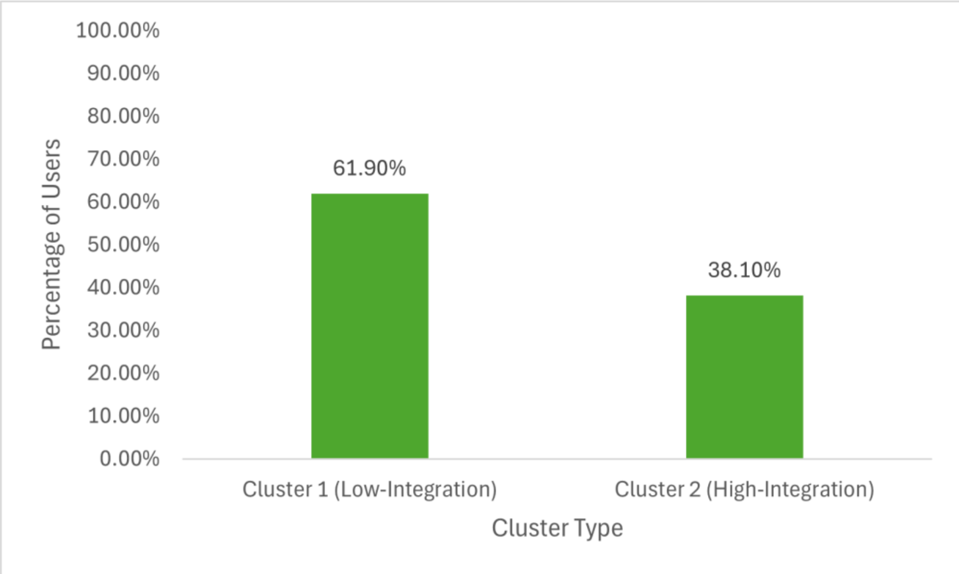
Silhouette analysis further supports this observation. The silhouette scores were 0.203 for $k = 2$, 0.164 for $k = 3$, 0.126 for $k = 4$, 0.134 for $k = 5$, and 0.085 for $k = 6$. The highest score was obtained for $k = 2$, indicating that this configuration provides the best balance between intra-cluster cohesion and inter-cluster separation.

Based on these results, the two-cluster solution was selected as the most appropriate representation of the dataset. This pattern is consistent with recent clustering-based segmentation studies that identified discrete user profiles rather than continuous usage gradients (Wedel & Kamakura, 2020; Everitt et al., 2021; Abdi & Williams, 2010).

Cluster Distribution

The final clustering model partitions the dataset into two distinct user groups. The first cluster contains 39 participants (61.9%), while the second cluster contains 24 participants (38.1%). This distribution indicates that the user population is not evenly segmented but instead dominated by a larger group of users with relatively lower levels of AI integration, alongside a substantial group of more active users. As illustrated in (Fig. 4), the majority of users belong to Cluster 1 (Low-Integration), while Cluster 2 (High-Integration) represents a smaller segment.

Fig. 4. Distribution of users across the identified clusters. Cluster 1 (Low-Integration Users) comprises 61.9% of participants, while Cluster 2 (High-Integration Users) accounts for 38.1%.



Cluster Profiles

Cluster 1: Low-Integration Users

The first cluster is characterized by relatively limited integration of AI into professional workflows. Users in this group exhibit moderate engagement with AI in personal contexts but significantly lower usage in professional settings. The mean personal AI usage score for this cluster is 0.724, while the mean professional usage score is 0.389. The average reported frequency of AI usage is 3.103, indicating occasional to moderate use.

In addition, users in this cluster report higher perceived barriers, with an average barrier-related score of 1.333. These barriers include concerns related to technical skills, trust in AI outputs, and privacy. Collectively, these characteristics suggest that users in this cluster tend to engage with AI selectively and have not yet fully integrated it into their daily professional activities.

A closer examination of barrier-related variables suggests that the most influential constraints are associated with limited technical proficiency and reduced trust in AI-generated outputs. These factors appear to restrict users' willingness to integrate AI into professional workflows, even when basic familiarity with AI tools is present. Privacy concerns also contribute to this hesitation, although to a lesser extent compared to skill and trust-related barriers. This indicates that improving user training and enhancing transparency in AI systems may play a critical role in facilitating the transition from low to high integration (Dwivedi et al., 2023; Kergroach, 2025; Capel & Brereton, 2023).

Volume 7, Issue 26 (2026) p428- 449

Cluster 2: High-Integration Users

The second cluster represents users with a significantly higher level of engagement with AI technologies. This group demonstrates strong usage across both personal and professional contexts, with a mean personal usage score of 1.612 and a mean professional usage score of 1.385. The average frequency of AI usage in this cluster is 4.083, indicating frequent interaction with AI tools.

Compared to Cluster 1, users in this group report slightly lower perceived barriers, with a mean value of 1.208. More importantly, their usage patterns indicate a shift toward task-oriented and operational applications of AI, including content preparation, evaluation tasks, and decision support. These characteristics suggest that AI is integrated into their workflows as a functional tool rather than used occasionally.

Table 1 summarizes the key characteristics of the identified clusters across the main feature dimensions.

Table 1. Summary of key characteristics across identified user clusters.

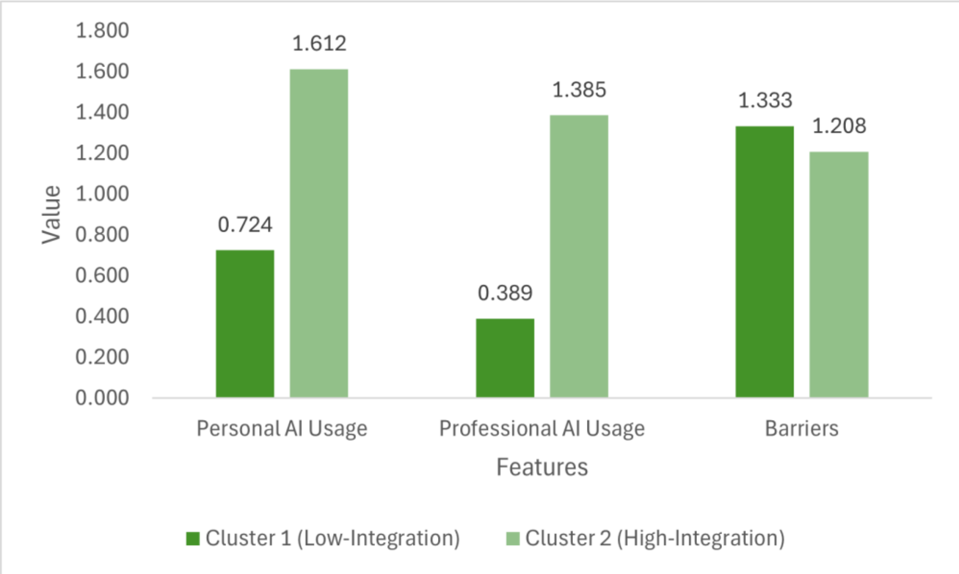
Feature	Cluster 1 (Low Integration)	Cluster 2 (High Integration)
Personal AI Usage	0.724	1.612
Professional AI Usage	0.389	1.385
AI Usage Frequency	3.103	4.083
Perceived Barriers	1.333	1.208

Comparative Analysis of Clusters

As shown in Table 1, the two clusters differ significantly in terms of professional AI usage and perceived barriers. A comparative analysis of the two clusters reveals that the primary distinction lies in the degree of professional integration and task-oriented usage of AI. While both clusters demonstrate some level of engagement with AI in personal contexts, the divergence becomes pronounced in professional applications (Rousseeuw, 1987; Hastie et al., 2021; Everitt et al., 2021; Dwivedi et al., 2023).

(Fig. 5) highlights significant differences between clusters in terms of personal AI usage, professional AI usage, and perceived barriers.

Fig. 5. Comparison of key characteristics across clusters. High-Integration users exhibit higher levels of AI usage (both personal and professional) and lower perceived barriers compared to Low-Integration users.



The most discriminative variables between clusters are associated with practical and applied use cases. These include the use of AI for preparing educational content, evaluating trainees, generating assignments, and supporting decision-making processes. In contrast, general usage activities such as casual information retrieval or basic content generation contribute less to cluster separation.

This finding indicates that the transition from low to high AI adoption is not simply a matter of increased usage frequency but involves a qualitative shift toward functional integration into domain-specific tasks. Users in the high-integration cluster utilize AI as an operational component of their workflows, whereas users in the low-integration cluster engage with AI in a more exploratory or limited manner.

Dimensionality Reduction Analysis

Principal component analysis (PCA) was applied to further investigate the structure of the dataset. The results indicate that a small number of principal components capture a substantial portion of the total variance, suggesting that user behavior can be effectively represented in a reduced feature space.

The leading components are primarily associated with variables related to professional and task-oriented AI usage, reinforcing the observation that these factors play a central role in differentiating user groups. This dimensionality reduction supports the validity of the clustering results and provides additional evidence that user segmentation is driven by a limited set of dominant behavioral dimensions.

This observation aligns with the clustering results, suggesting that the separation between clusters is

Volume 7, Issue 26 (2026) p428- 449

largely driven by these dominant components, particularly those associated with professional and task-oriented AI usage, thereby reinforcing the validity of the identified user segmentation (Han et al., 2011; Abdi & Williams, 2010).

Implications for User Modeling

The identified cluster structure provides a clear basis for computational user modeling. The two clusters correspond to distinct behavioral profiles that can be directly mapped to system-level representations. The low-integration users can be characterized by moderate personal usage, limited professional integration, and higher perceived barriers, while the high-integration users exhibit consistent and task-oriented usage across contexts.

This segmentation enables the construction of compact and interpretable user models that make user behavior patterns more transparent and explainable for system designers (Zhang & Chen, 2020; Adadi & Berrada, 2018). By representing users in terms of usage intensity, professional integration, and barrier levels, systems can adapt their behavior to better match user needs and capabilities.

Discussion

The results of the clustering analysis provide several important insights into the structure of AI usage behavior in professional environments. These findings not only reveal the structural segmentation of users but also provide a deeper understanding of the behavioral mechanisms that drive differences in AI integration across professional contexts. Most notably, the data reveals a bifurcated user structure, where participants are naturally divided into two dominant groups rather than forming a gradual spectrum of adoption. This finding suggests that AI usage is not best understood as a linear progression from low to high engagement, but rather as a structured phenomenon characterized by distinct behavioral modes. As illustrated in the proposed framework (Fig. 1), this bifurcated structure supports the transition from data-driven clustering to system-level personalization, where distinct user profiles inform adaptive system behavior (Brusilovsky, 2001; Gauch et al., 2007; Hastie et al., 2021).

From a data mining perspective, the emergence of a two-cluster solution as the optimal configuration, supported by both elbow analysis and silhouette scores-suggests that the underlying feature space may be largely organized around a dominant dimension. This dimension reflects the degree of functional integration of AI into professional tasks, rather than general familiarity or exposure. In other words, what differentiates users is not whether they use AI, but how they use it.

The identified clusters are interpretable in terms of observable behavioral features, aligning with principles of explainable artificial intelligence that emphasize transparency in model outputs (Adadi & Berrada, 2018; Zhang & Chen, 2020).

The distinction between clusters highlights a critical shift from exploratory usage to operational usage. Users in the low-integration cluster engage with AI in a limited and often context-independent manner, typically for basic or occasional tasks. In contrast, users in the high-integration cluster incorporate AI into structured workflows, applying it to domain-specific activities such as content preparation, evaluation, and decision support. This transition represents a qualitative change in the role of AI, from a supplementary tool to an integral component of professional practice.

These findings are consistent with recent research in user modeling and personalization, which emphasize that user populations often exhibit heterogeneous and sometimes polarized behavior when interacting with emerging technologies. Clustering-based approaches have been shown to effectively capture such heterogeneity, enabling the identification of user groups that can support adaptive system design. The present study extends this line of work by demonstrating that similar patterns can be identified using structured survey data, rather than relying solely on system-generated interaction logs (Brusilovsky, 2001; Amershi et al., 2019; Raisch & Krakowski, 2021; Wedel & Kamakura, 2020).

Another important observation is the role of task-oriented variables in driving cluster separation. The most discriminative features in the dataset are associated with applied use cases, particularly those related to professional activities. This suggests that adoption is influenced less by general attitudes toward AI and more by the extent to which users perceive AI as relevant and useful within their specific domain tasks. Consequently, efforts to increase AI adoption should focus not only on improving accessibility or awareness but also on enhancing the alignment between AI capabilities and domain-specific needs (Rousseeuw, 1987; Kergroach, 2025; OECD, 2025; Ricci et al., 2022).

The use of principal component analysis further supports these conclusions by showing that a limited number of components explain a significant portion of the variance in the dataset. This indicates that user behavior can be effectively represented using a reduced set of dimensions, which aligns with the proposed user profiling model. From a computational standpoint, this finding is important, as it suggests that efficient and scalable user models can be constructed without requiring high-dimensional representations (Han et al., 2011; Abdi & Williams, 2010).

From the perspective of intelligent system design, the identified clusters provide a practical foundation for adaptive and personalized behavior. Systems can differentiate between user types and adjust functionality accordingly. For selective users, systems may prioritize simplicity, guidance, and trust-building features, reducing barriers to adoption. For active users, systems can offer advanced capabilities, automation, and customization, leveraging their familiarity with AI tools. This approach enables systems to move beyond static design toward dynamic, user-aware interaction models (Kobsa, 2007; Jannach et al., 2021; Zhang & Chen, 2020).

In terms of contribution, this study demonstrates that clustering techniques can be effectively applied to multidimensional survey data to produce validated and interpretable user profiles. By integrating clustering, validation, and dimensionality reduction, the proposed framework provides a robust methodology for user modeling in AI-enabled environments. Unlike many prior studies that rely on predictive modeling, this work emphasizes exploratory analysis and the discovery of latent structures, offering a complementary perspective on user behavior.

However, several limitations should be acknowledged. The dataset used in this study comprises 62 participants, which, while adequate for exploratory analysis and methodological validation, may limit the generalizability of the findings to broader populations. The relatively small sample size may also restrict the diversity of behavioral patterns captured, potentially affecting the stability of the identified clustering structure across different contexts. Future research should evaluate the proposed framework using larger and more heterogeneous datasets to assess its scalability and robustness across diverse user populations and professional environments (OECD, 2025; Tarafdar et al., 2022; Kergroach, 2025; OECD, BCG, & INSEAD, 2025).

Overall, the findings highlight the importance of data-driven user segmentation in understanding AI adoption and designing intelligent systems. By uncovering latent behavioral patterns, the proposed approach contributes to the development of adaptive systems that are better aligned with user needs and usage contexts. This highlights the importance of designing AI systems that are not only technically capable but also aligned with users' operational needs and readiness levels.

Conclusion

This study presented a data-driven framework for profiling AI users in professional environments using unsupervised learning techniques. By applying a hybrid clustering approach to a multidimensional dataset, the analysis identified a stable two-cluster structure representing distinct patterns of AI usage behavior. The results demonstrate that AI adoption is not a gradual continuum but is better characterized by a clear separation between selective users with limited professional integration and active users who incorporate AI into task-oriented workflows.

From a methodological perspective, the study highlights the effectiveness of combining clustering, validation techniques, and dimensionality reduction to derive robust and interpretable user profiles. The use of both K-means and hierarchical clustering, along with silhouette-based validation, ensures that the identified segmentation reflects meaningful structures in the data rather than artifacts of a single algorithm. Furthermore, the application of principal component analysis demonstrates that user behavior can be represented using a reduced set of dominant features, supporting efficient and scalable modeling.

From a systems perspective, the findings provide a foundation for adaptive and personalized AI system design. The identified user profiles enable systems to differentiate between levels of AI integration and adjust functionality accordingly, supporting more effective interaction strategies and improved user experience. By linking user segmentation to system-level adaptation, the proposed framework contributes to the development of intelligent systems that are responsive to diverse user needs (Kobsa, 2007; Jannach et al., 2021; Zhang & Chen, 2020; Gauch et al., 2007).

In addition, the study demonstrates that structured survey data can be effectively utilized as input for data mining approaches, extending beyond traditional reliance on system-generated interaction data. This expands the range of data sources available for user modeling and opens new opportunities for integrating behavioral and perceptual dimensions into computational frameworks.

Future research may extend this work by incorporating larger and more diverse datasets, exploring alternative clustering techniques, and evaluating the integration of user profiles within real-world AI systems. In addition, a promising direction for future investigation is the development of longitudinal user profiling approaches, where user behavior is tracked over time to analyze transitions between clusters. Recent work on adoption personas and large-scale survey-based AI profiling suggests that such transitions can be examined meaningfully over time and across domains (Wedel & Kamakura, 2020; Ricci et al., 2022; Everitt et al., 2021;

Capel & Brereton, 2023). Such an approach would enable the examination of how users evolve from low-integration to high-integration profiles as they gain experience, improve technical skills, and develop greater trust in AI systems. This dynamic perspective could further enhance the design of adaptive systems by supporting personalized interventions that facilitate user progression across integration levels.

Disclosure Statements:

- **Ethical approval and consent to participate:** Participation in the research was approved in accordance with the journal's guidelines.
- **Availability of data and materials:** All data and materials are available upon request.
- **Authors' contributions:** The authors are responsible for all aspects of the research, including content, analysis, methodology, and the final review.
- **Conflicts of interest:** The authors declare that there are no conflicts of interest related to the design, submission, or evaluation of this research.
- **Funding:** This research received no specific funding.
- **Acknowledgements:** The authors would like to express their sincere appreciation to Journal of Scientific Conferences (JSC) for its support and guidance (https://jsd.sdasmart.org).

Bibliography List:

Abdi, H., & Williams, L. J. (2010). Principal component analysis. Wiley Interdisciplinary Reviews: Computational Statistics, 2(4), 433–459.

Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). IEEE Access, 6, 52138–52160.

Alannsary, M. (2025). Acceptance of the Use of Artificial Intelligence Applications and its role in Developing Human Resources Among Faculty Members: An Applied Study at a Training Institution in the Kingdom of Saudi Arabia. Journal of Scientific Development for Studies and Research (JSD). Vol. 6, Issue 23. pp 177-207

Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., ... Horvitz, E. (2019). Guidelines for human-AI interaction. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems.

Bishop, C. M. (2006). Pattern recognition and machine learning. Springer.

Brusilovsky, P. (2001). Adaptive hypermedia. User Modeling and User-Adapted Interaction, 11(1–2), 87–110.

Capel, T., & Brereton, M. (2023). What is human-centered AI? Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems.

Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., ... Williams, M. D. (2023). Artificial intelligence (AI): Multidisciplinary perspectives on emerging challenges. International Journal of Information Management, 71, 102642.

Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2021). Cluster analysis (5th ed.). Wiley.

Fischer, G. (2001). User modeling in human–computer interaction. User Modeling and User-Adapted Interaction, 11(1), 65–86.

Gauch, S., Speretta, M., Chandramouli, A., & Micarelli, A. (2007). User profiles for personalized information access. In The adaptive web (pp. 54–89). Springer.

Géron, A. (2022). Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow (2nd ed.). O'Reilly.

Han, J., Kamber, M., & Pei, J. (2011). Data mining: Concepts and techniques (3rd ed.). Elsevier.

Hastie, T., Tibshirani, R., & Friedman, J. (2021). The elements of statistical learning (2nd ed.). Springer.

Jain, A. K. (2010). Data clustering: 50 years beyond K-means. Pattern Recognition Letters, 31(8), 651–666.

Jannach, D., Zanker, M., Felfernig, A., & Friedrich, G. (2021). Recommender systems: An introduction. Cambridge University Press.

Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. Philosophical Transactions of the Royal Society A, 374(2065), 20150202.

Kaufman, L., & Rousseeuw, P. J. (2020). Finding groups in data: An introduction to cluster analysis. Wiley.

Kergroach, S. (2025). Artificial intelligence adoption in small and medium-sized enterprises. OECD Digital Economy Papers.

Knijnenburg, B. P., & Willemsen, M. C. (2020). Evaluating recommender systems with user-centric metrics. User Modeling and User-Adapted Interaction, 30, 1–36.

Kobsa, A. (2007). Generic user modeling systems. In The adaptive web (pp. 136–154). Springer.

Lloyd, S. (1982). Least squares quantization in PCM. IEEE Transactions on Information Theory, 28(2), 129–137.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. Proceedings of the Fifth Berkeley Symposium.

Murphy, K. P. (2022). Probabilistic machine learning: An introduction. MIT Press.

OECD. (2025). Bridging the AI divide. OECD Publishing.

OECD, Boston Consulting Group, & INSEAD. (2025). The adoption of artificial intelligence in firms. OECD Publishing.

Raisch, S., & Krakowski, S. (2021). Artificial intelligence and management. Academy of Management Review, 46(1), 192–210.

Ricci, F., Rokach, L., & Shapira, B. (2022). Recommender systems handbook (3rd ed.). Springer.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation of cluster analysis. Journal of Computational and Applied Mathematics, 20, 53–65.

Saudi Data and Artificial Intelligence Authority [SDAIA]. (2024). State of artificial intelligence in Saudi Arabia. Riyadh, Saudi Arabia: SDAIA. Retrieved from https://sdaia.gov.sa/en/MediaCenter/KnowledgeCenter/ResearchLibrary/StateofAIinSaudiArabia.pdf

Shlens, J. (2014). A tutorial on principal component analysis. arXiv:1404.1100.

Tarafdar, M., Beath, C. M., & Ross, J. W. (2022). Using AI to enhance business operations. MIT Sloan Management Review.

Venkatesh, V., Morris, M., Davis, G., & Davis, F. (2003). User acceptance of information technology. MIS Quarterly, 27(3), 425–478.

Wedel, M., & Kamakura, W. (2020). Market segmentation: Conceptual and methodological foundations. Springer.

Zhang, Y., & Chen, X. (2020). Explainable recommendation: A survey. IEEE Access, 8, 102194–102217.