International Journal of Basic and Applied Sciences, 14 (SI-3) (2025) 63-72

International Journal of Basic and Applied Sciences

Website: www.sciencepubco.com/index.php/IJBAS
https://doi.org/10.14419/txmrrk92
Research paper

Multimodal Content Analysis for Enhanced User Experience

Xiang Li *, Tamprasirt Anukul, Fangli Ying

International College of Digital Innovation, Chiang Mai University, Chiang Mai, Thailand
**Corresponding author E-mail:xiang325.li@gmail.com*

Received: April 29, 2025, Accepted: July 22, 2025, Published: August 15, 2025

Abstract

The aim of the study was to develop methods of multimodal content analysis to improve the user experience in interactive systems. The study reviewed existing approaches to multimodal content analysis, established criteria for developing new methods, and provided specific examples of practical application of the developed methods in various fields, demonstrating their effectiveness and potential in real-world conditions. The main results consisted of the development of methods, in particular, integration and synchronization of modalities, which can demonstrate high efficiency in medical diagnostics. In turn, the personalization of user experience in streaming services increases user satisfaction through relevant recommendations, while ensuring data privacy and security meet modern regulatory requirements in the healthcare sector. The paper also examines the integration of modalities, which includes convolutional neural networks for image and video processing, recurrent neural networks for text and audio processing, and attention mechanisms for highlighting important parts of the data. Multimodal analysis involves the processing and integration of data from different sources, which provides a complete picture to improve the analysis. The developed methods outperform most existing approaches, providing higher accuracy, speed, robustness to noise, and incomplete data.

Keywords: *Improving the Quality of the Platform; Interaction of Participants; Interactive Systems; Multimodal Review; Study of Materials.*

1. Introduction

In the contemporary digital landscape, interactive systems play a central role in organizing, processing, and delivering information across a wide array of domains. However, the growing complexity and volume of multimodal data, including text, audio, visual, and behavioral inputs, highlight the limitations of existing analytical frameworks in terms of accuracy, processing speed, and adaptability. These limitations hinder the development of user-oriented services that are capable of dynamic personalization, robust performance in noisy environments, and secure handling of sensitive data.

This study addresses the pressing need to enhance multimodal content analysis methods that support efficient data integration and improve user experience in interactive systems. To identify methodological gaps and application constraints, we analyzed a selection of recent studies relevant to multimodal systems, interface usability, personalization techniques, and data security. These include works by Chaiyaphum (2022), Yasinthorn and Celadon (2023), Kamtab and Wiriyanon (2022), Ramtohul and Khedo (2023), Yasmin and Jamal (2023), Fang et al. (2024), Jagnade et al. (2023), Heraz et al. (2024), Dewangan and Mishra (2023), and Alzubi et al. (2023), which were selected for their focus on interface design, multimodal integration, system personalization, and user-centered performance metrics.

This article aims to develop, systematize, and evaluate innovative methods of multimodal content analysis to enhance the user experience in various interactive contexts. The objectives of the study include reviewing existing approaches to multimodal content analysis, evaluating the developed methods, comparing the results with existing approaches, and identifying the advantages of new methods.

2. Literature Review

Numerous studies have explored the potential of multimodal approaches to improve user interaction and system responsiveness across various contexts. For example, the results of the study by Chaiyaphum (2022) showed that the development of innovative approaches in the supply chain and the adaptation of operational processes in logistics can significantly affect the improvement of interaction between different stages. However, some issues require further study, in particular, the adaptation of these approaches to ensure improved user experience. The authors Yasinthorn & Celadon (2023) demonstrated the use of WordPress for website development with a focus on computer technology and cybersecurity, which contributed to improved security and resource optimization. However, the study did not assess the impact of the approaches used on user experience and overall user satisfaction, which is an important aspect for further research.

Researchers Kamtab & Wiriyanon (2022) provided results on assessing the quality of design and testing the effectiveness of a web application that are of the highest quality. Nevertheless, there are problems such as the need to adapt the developed approaches to improve the effectiveness of user interaction, especially in the context of long-term use by elderly users. And the results of the work of Yasmin & Jamal (2023) have shown that a multimodal approach is more effective in improving the overall user experience and engagement compared to

unimodal systems based on visual perception alone. The problem is that multimodality requires more complex integration and optimization to ensure continuous and high-quality operation, especially when interacting with large amounts of data or in virtual reality. In addition, Fang et al. (2024) pointed out that combining reviews with multimodal data significantly improves the accuracy of recommendations and the overall user experience. However, there is a problem of additional noise that occurs when direct integration of reviews occurs. In addition, the results of Jagnade et al. (2023) demonstrated that the integration of multiple modalities significantly improves the accuracy of cursor control and reduces cognitive load. However, the problem is the need to further study the adaptation of the system to individual user preferences and environmental conditions. Additionally, Ramtohul & Khedo (2023) showed that partial implementation of the user experience framework significantly improved user engagement in prototype systems. However, further research is needed to fully implement the user experience framework and adapt it to a variety of applications and user needs.

In turn, the results of the study by Heraz et al. (2024) indicated that tracking emotions using gestures can increase user engagement, and the best results were achieved using the LightGBM algorithm, which demonstrated very high accuracy. At the same time, certain issues require further study, such as the adaptation of gesture technologies to improve user interaction in a multimodal environment. Moreover, Dewangan & Mishra (2023) proposed an effective transfer learning model with autoencoders that achieves very high accuracy compared to state-of-the-art sentiment analysis methods. However, the integration of multimodal data to improve user experience in real-world settings requires further study. Finally, researchers & Alzubi et al. (2023) developed a multimodal deep learning multilayer convolutional neural network to improve the educational experience, and the results showed high real-time performance, including high-quality interaction and improved educational services. Nevertheless, there are challenges such as integrating multimodal approaches to improve the user experience in learning environments.

These findings collectively underline the necessity of advancing multimodal analysis methods that are not only technically robust but also capable of enhancing individual user experience. The current study builds on these works by proposing integrated, secure, personalised, and computationally optimised methods suitable for a wide range of domains, including healthcare, streaming services, and educational platforms.

3. Materials and methods

This study employed a mixed-methods research design that integrated theoretical modeling with empirical evaluation to develop and validate new methods of multimodal content analysis aimed at enhancing user experience in interactive systems. A review of existing approaches, for example, Yasmin & Jamal (2023), Fang et al. (2024), Jagnade et al. (2023), Alzubi et al. (2023), helped to identify their advantages, disadvantages, and practical applications. Various methods of data processing and integration, such as text, audio, video, and image analysis, were discussed. The study of practical examples of multimodal content application helped to identify key requirements for systems and algorithms to achieve high efficiency and accuracy of analysis.

Identifying criteria for the development of new methods was one of the main tasks of this study. Based on these criteria, methods for multimodal content analysis were developed. Integration and synchronization of modalities involved combining data from different sources into a single stream of information. For this purpose, signal processing and machine learning were used to synchronize text, audio, video, and other types of data. This method not only ensures high-quality integration but also guarantees the reliability and availability of all information for further analysis and processing. Also, important was the optimization of algorithms for computational efficiency, which included reducing the computational complexity of algorithms and reducing resource consumption using boosting and reducing the dimensionality of data, such as auto-encoders. This significantly reduced computational costs and increased processing speed, which played a critical role in ensuring the prompt and efficient integration of various types of data into a single information flow. Personalization of the user experience involved the development of schemes that adapt the system to the individual needs of users. Clustering and recommender systems were studied to provide personalized recommendations. Thus, personalization schemes were developed using clustering methods to identify user preferences, which can help improve the quality of service and ensure an individual approach to each user of the system.

Ensuring data confidentiality and security was important, as it included the implementation of data protection mechanisms that meet modern regulatory requirements. For this purpose, data encryption and anonymization methods were used to guarantee compliance with confidentiality requirements and ensure the security of personal and confidential data of system users. To increase robustness to noise and incomplete data, regression and clustering methods were used to reduce the impact of incorrect or missing data. This was achieved by using statistical analysis and machine learning techniques that allow effective modelling of data dependencies and making reliable predictions even in difficult conditions, such as high noise or insufficient information.

The developed methods were also compared with existing approaches, which showed the advantages of the new methods in terms of accuracy and data processing speed. This approach allowed creating a basis for further development and optimization of new methods of multimodal content analysis aimed at improving user experience and data processing efficiency.

4. Results

4.1 Theoretical foundations and overview of existing approaches to multimodal content analysis

This approach involves integrating and analysing data from various sources and types, such as text, images, video, and audio, to gain a deeper and more multifaceted understanding of information. Multimodal analytics can significantly improve the user experience, as it enables the creation of more interactive, personalized, and relevant services and applications.

The basic idea behind multimodal analysis is that each type of data provides unique information that complements the others. For example, textual data can contain specific facts and descriptions, while images can provide visual information, and video and audio can provide dynamic context. The combination of these different modalities allows for a more complete picture and enhanced analysis capabilities. Multimodal content analysis is used in various industries, including medicine, education, marketing, entertainment, and many others (Table 1).

Table 1: Examples of multimodal analysis in different industries

Industry	Example of use
Medicine	Combining text records, medical images, and test results to improve diagnosis
Education	Creating interactive learning materials that include text, video, and audio
Marketing	Analysing user reviews, product images, and promotional videos to improve marketing strategies
Entertainment.	Integrate text, audio, and video to create more immersive games and films
Finance.	Analysing financial reports using text, graphs, and audio commentary to predict market trends
Tourism	Combining text descriptions, images of tourist attractions, and video tours to create personalized itineraries
Technologies	Analysing technical documents, diagrams, and video instructions to improve technical support and training

Source: created by the author based on Chiu et al. (2024), Bouchey et al. (2021), Rust (2020), Meier (2022), Henrique (2023), Marques & Marques (2023), Naegelin et al. (2023).

One of the key advantages of multimodal analysis is the ability to obtain more accurate and relevant results by using additional information from different sources. This is especially important in cases where data from one source may be incomplete or inaccurate. On the other hand, multimodal analysis also presents significant challenges related to the integration and synchronization of data from different modalities, as well as the need for high computing power and complex data processing algorithms. It is also worth highlighting the general process of multimodal analysis, including data integration, complementarity of modalities and information multiplexing (Fig. 1).

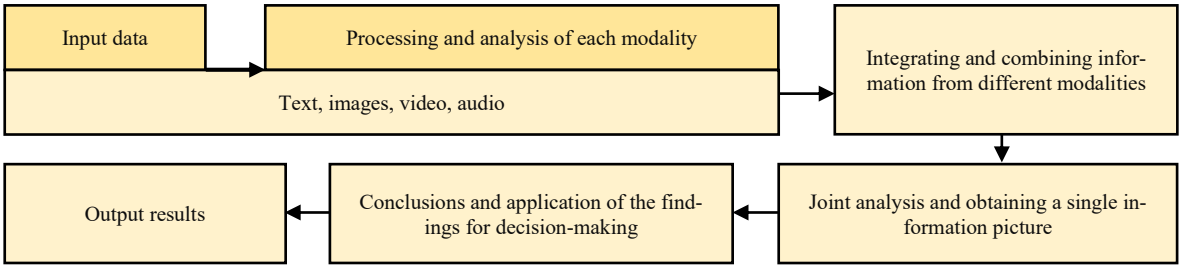

Fig. 1: The process of multimodal analysis

Source: created by the author based on Wildfeuer & Lehmann (2024).

The key aspects of the theoretical foundations of multimodal analysis are revealed through several important aspects that help to understand the essence of this approach. The synergistic effect is one of the fundamental aspects of multimodal analysis (Khoda et al., 2024; Maksymov & Tryus, 2025). The combination of different modalities creates a unique synergistic effect, which allows for more total information than could be obtained by analysing each modality separately.

Computer vision is another key aspect that plays an important role in multimodal analysis (Khrulov, 2025; Zhetenbayev et al., 2022). The use of computer algorithms to process, analyse, and integrate data from different sources is critical to the successful implementation of this approach. This includes not only the technical aspects of data processing but also the development of algorithms that can effectively work with different types and formats of information, ensuring its integration and analysis in a single information system.

Modelling the interaction of modalities is another important component of the theoretical foundations of multimodal analysis. The development of models that consider the interaction of different modalities allows not only combining data, but also using it to achieve specific analytical goals (Kerimkhulle et al., 2023; Pavlova et al., 2024). This may include the development of data processing methods that consider the specific relationships between text, image, video, and audio, which improves the quality and accuracy of the analysis. These aspects together form the basis for the implementation of multimodal analysis, which opens new opportunities for understanding, analysing, and using a variety of information from different sources and formats. Moreover, it is necessary to distinguish modern approaches to multimodal analysis, which can be divided into three main categories: early fusion, late fusion, and deep learning (Table 2).

Table 2: Comparison of approaches to multimodal analysis

	Early merger	Late merger	Deep learning
Advantages	Efficiency in detecting coarse correlations between data Ease of implementation	Ability to save information about the data structure Greater resistance to noise	Power to explore complex relationships between data High precision and adaptability
Disadvantages	Sensitivity to noise Differences in data scale can skew results	The possibility of losing important information about the relationships between data	Complexity of implementation High computing costs
Example of use	Combining text and images to improve image search results	Analysing text and audio to determine sentiment based on non-verbal cues	Using co-bedding to integrate text and images in image description generation models

Source: created by the author based on Taylor et al. (2021).

In early fusion, different types of data are merged at an early stage of analysis. This can be done through data concatenation, addition, or integration. In late fusion, different types of data are analysed separately, and then the results are fused at a later stage, which can be done through comparison, matching, or multiple inference. Deep learning employs convolutional neural networks for image processing and recurrent neural networks for sequential data analysis. This method is used through multitasking, joint embedding, or attention.

Multimodal data analysis has a wide range of applications. For example, in object recognition, the combination of information from images, text, and audio can significantly improve the accuracy of object detection. This provides more reliable results than analysing only one type of data. In summary, multimodal analysis is a progressive field with a wide range of potential applications, so expect even more innovative applications of this technology in the future.

4.2 Developing methods to improve user experience based on multimodal analysis

Developing methods for improving user experience based on multimodal analysis requires careful definition of criteria that will ensure the effectiveness and relevance of these methods. One of the main criteria is the ability of methods to effectively integrate different data modalities. Each modality provides unique information, and their combination should create a synergistic effect that allows for more

complete and accurate information. To ensure the correct integration of different modalities, the data must be synchronized in time and context. This means that methods must consider the temporal and spatial dependencies between different types of data to ensure that they are combined correctly.

Methods should ensure high accuracy of analysis and relevance of results, which is especially important for decision-making applications such as medical diagnostics or recommendation systems. The results should not only be accurate, but also useful and understandable to end users. Since multimodal analysis can be computationally intensive, developing methods that are computationally efficient is crucial. This includes optimizing algorithms to process large amounts of data quickly and efficiently, including using machine learning and deep learning techniques. In addition, methods must be scalable to be able to handle large and heterogeneous data sets. This includes the ability to process data in real time, as well as to adapt to growing data volumes and user requirements. In real-world settings, data may be noisy or incomplete (Orazbayev et al., 2023; Xhafka et al., 2024). Therefore, methods need to be robust to such challenges, i.e., able to work effectively even in the presence of inaccurate or missing data.

To improve the user experience, it is essential to consider the individual characteristics of users (Iklassova et al., 2024; Kovalchuk, 2025). Methods should support the personalization of content and services by adapting to users' interests, preferences, and behaviours. In addition, methods should be designed to be user-friendly for end users. This includes an intuitive interface, quick access to the necessary information, and support for different platforms and devices. When developing new methods, data privacy and security should also be considered. Methods should protect users' personal information and comply with data security standards.

Consideration of these criteria is important for creating effective and reliable methods for multimodal analysis that will improve the user experience in various industries. The first method is the integration and synchronization of modalities (Fig. 2). Its purpose is to ensure that different types of data are effectively combined and synchronized to create a synergistic effect. This method involves the use of algorithms to integrate text, images, audio, and video. The system first preprocesses the data of each modality, considering temporal and spatial dependencies. Then the data is synchronized, which allows it to be combined correctly. For this purpose, the system uses a joint embedding technique that transforms data from different modalities into a common feature space, which facilitates their integration.

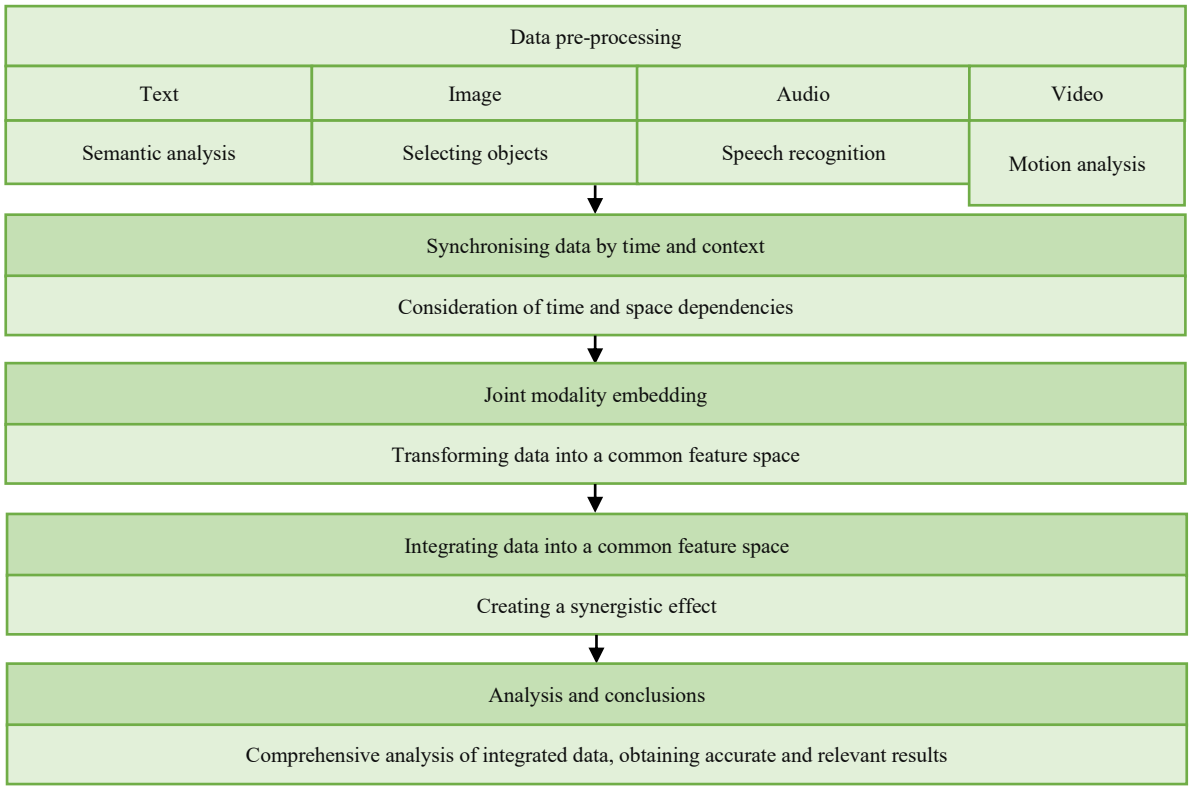

Fig. 2: The process of integrating and synchronizing modalities

Source: created by the author based on Ziker et al. (2021).

An important method is to optimize algorithms for computational efficiency (Fig. 3). Its purpose is to develop approaches to ensure fast and efficient processing of large volumes of multimodal data. This method involves the use of optimized machine learning and deep learning algorithms. It is based on the principle of reducing computational complexity by simplifying models and using more efficient data processing methods, such as quantization and network pruning. Quantization is the process of reducing the number of bits required to represent numbers in a machine learning model, while pruning is the process of removing redundant or less important parts of a neural network to reduce its size and increase efficiency. In addition, parallel data processing and distributed computing are used to reduce processing time.

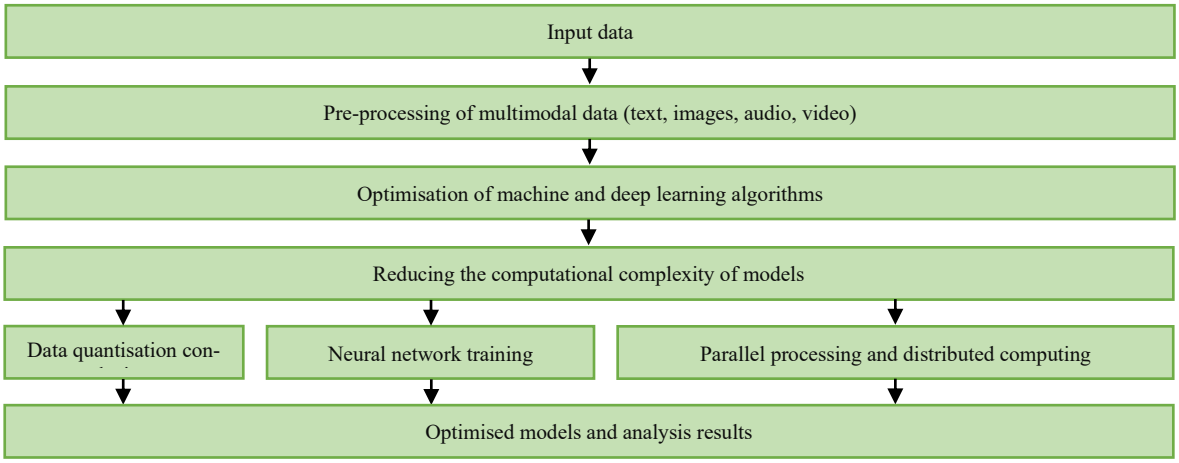

Fig. 3: Optimizing algorithms for computational efficiency

Source: created by the author based on Mao & Zhang (2021).

It is also worth paying attention to the method of robustness to noise and incomplete data (Fig. 4). The purpose of this method is to ensure the robustness of various methods when processing data with noise or missing values. It involves the use of techniques to fill in missing data and filter out noise. A combination of data imputation and model regularization is used. Imputation is the process of replacing missing or omitted values in a dataset with certain estimates based on algorithms that consider contextual information, while regularization helps to avoid overfitting on noisy data. The use of data augmentation helps to increase the resilience of models to data variability. Augmentation is the process of artificially increasing the amount of data by modifying it or creating new variants based on existing data.

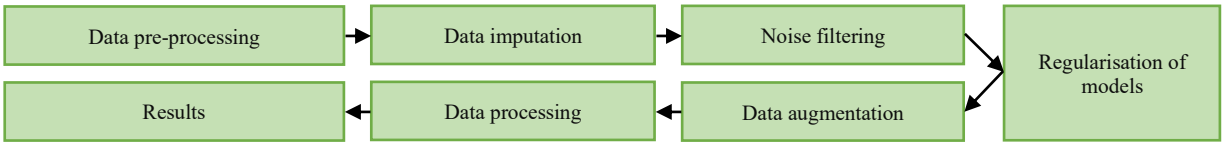

Fig. 4: Robustness to noise and incomplete data

Source: created by the author based on Ragan & Villarin (2021).

In turn, attention should be focused on the method of personalizing user experience, which aims to develop ways to adapt content and services to the individual needs of users (Fig. 5). This method involves the use of recommendation system algorithms that consider users' behavioural data, interests, and preferences. A hybrid approach is used that combines collaborative filtering and content analysis. This allows creating personalized recommendations that meet the unique needs of each user.

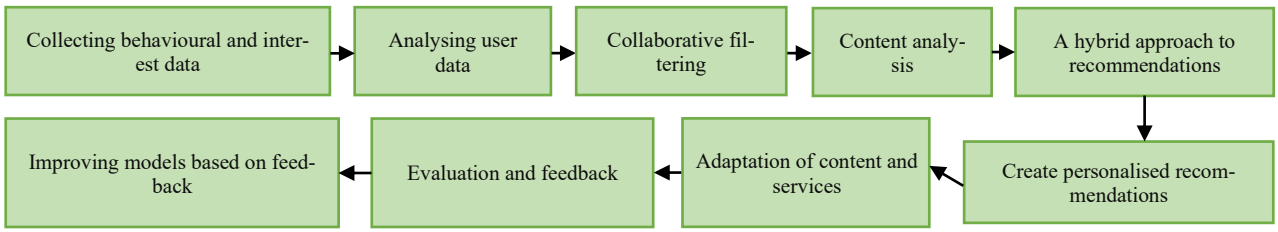

Fig. 5: Stages of user experience personalization

Source: created by the author based on UI/UX staffing (2019).

Moreover, a method of ensuring data privacy and security (Fig. 6) can be distinguished. Its purpose is to protect users' personal information and comply with data security standards. This method includes the implementation of cryptographic algorithms to protect data and comply with privacy standards. Data anonymization is used to protect personal information without losing the usefulness of the data for analysis. Anonymization is a data processing process aimed at removing or masking information that can be used to identify specific individuals. Additionally, differential privacy techniques are being implemented to ensure that data is protected even when it is analysed.

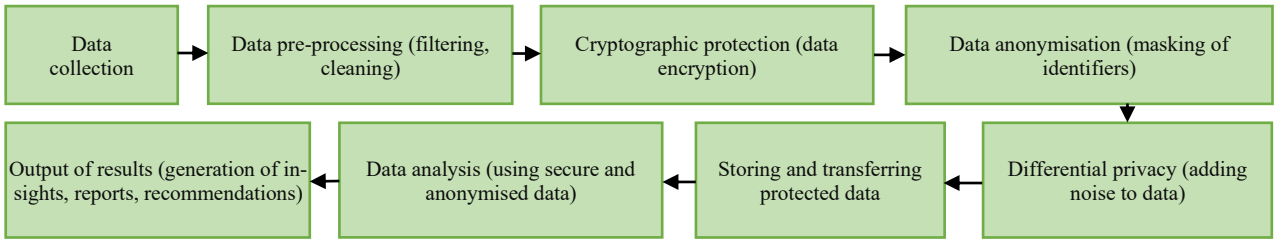

Fig. 6: Ensuring data privacy and security

Source: created by the author based on What is Data Protection and Privacy? (2024).

In general, the implementation of these methods will improve the accuracy and speed of analysis, adapting the results to the individual needs of users, while ensuring compliance with modern regulatory requirements in various fields.

4.3 Practical application and comparison of the developed methods

The developed methods for improving the user experience based on multimodal analysis have significant potential for implementation in various real-world applications and systems. Each of the discussed methods can be applied to solve specific problems in various industries. Modalities integration and synchronization can be effectively used in medical diagnostics, where combining data from different sources, such as textual notes from doctors, images from medical scans, and laboratory test results, is crucial. For example, a system that integrates data from different modalities can improve diagnostic accuracy and help doctors make decisions by providing a comprehensive view of a patient's condition.

The method of optimizing algorithms for computational efficiency can be used in large data centres and data processing systems where speed and efficiency are critical. For example, in financial institutions that process large volumes of transactional data, the use of optimized algorithms can significantly reduce computing resource costs and increase transaction processing speed. Also, the method of robustness to noise and incomplete data can be applied to environmental monitoring systems where data often contain noise or may be incomplete due to technical failures. The use of this method will increase the accuracy and reliability of environmental data analysis, providing more accurate forecasts and recommendations for reducing negative environmental impacts. In addition, personalization of user experience can be implemented in e-commerce systems and streaming services, where it is important to provide users with relevant recommendations. For example, in streaming services such as Netflix or Spotify, personalized recommendations can improve user satisfaction by offering content that matches their preferences and interests. This method differs from existing approaches in its ability to process and integrate more data modalities to create a more complete user profile. Unlike traditional systems, it considers the temporal and spatial dependencies between different types of data to provide more accurate and relevant recommendations and uses advanced machine learning techniques such as deep learning and co-embedding to create more complex and accurate models of user behaviour. As a result, this method can adapt to more diverse and specific user needs, which is particularly useful in dynamic environments. In turn, the method of ensuring data confidentiality and security is critical for all systems that work with personal user data. For example, in healthcare systems that store and process sensitive patient medical data, the use of anonymization and cryptographic protection methods will ensure confidentiality and compliance with regulatory requirements.

These methods for improving user experience based on multimodal analysis have certain advantages over existing approaches that are worth considering. For example, integration and synchronization of modalities include joint embedding, data synchronization, and integration of text, images, audio, and video. Optimizing algorithms for computational efficiency uses quantization, pruning, and parallel processing (Bezshyyko et al., 2008; Porkodi & Raman, 2025). Robustness to noise and incomplete data is achieved through data imputation, regularization, and augmentation (Aviv et al., 2023; Tokarieva et al., 2024). In addition, personalization of user experience includes a hybrid approach, namely, collaborative filtering and content analysis. As for data privacy and security, it includes cryptographic algorithms, anonymization, and differential privacy. In other words, these methods ensure efficient data processing, accurate analysis, increased resistance to noise and data omissions, personalized user experience, as well as a high level of data confidentiality and security. This allows creating more reliable systems that meet modern requirements for processing and analysing multimodal data.

To evaluate the effectiveness of these methods, it is worth considering accuracy, processing speed, resistance to noise and data omissions, level of personalization, and confidentiality. These methods demonstrate high accuracy of analysis due to the effective integration and synchronization of modalities, which allows for more comprehensive and accurate information and significantly improves the quality of the results. Optimized algorithms using quantization and pruning significantly reduce computational complexity, enabling fast processing of large amounts of data (Tkachenko et al., 2025; Kerimkhulle & Aitkozha, 2017). Parallel processing and distributed computing further reduce processing time. The use of data imputation, regularization, and augmentation techniques allows increasing the robustness of models to noise and missing data, which ensures stable and reliable results even in difficult conditions (Azieva et al., 2021; Rexhepi et al., 2023).

A hybrid approach to personalizing the user experience, combining collaborative filtering and content analysis, ensures a high level of relevance in recommendations. This helps to improve user satisfaction and increase engagement. The implementation of cryptographic algorithms and data anonymization ensures a high level of protection of users' personal information, meeting modern privacy and security standards (Kolodiziev et al., 2021; Konurbayeva et al., 2015). In addition, the efficient processing of various types of data has resulted in a more complete and accurate presentation of information. Reducing computational complexity significantly speeds up the process of analysing large amounts of data (Kerimkhulle et al., 2021; Bisenovna et al., 2024). The use of advanced data processing techniques ensures the reliability and stability of the results. A hybrid approach to recommendations allows tailoring content and services to the individual needs of users, increasing their satisfaction. The implementation of modern data protection methods guarantees a high level of security for users' personal information (Varanitskyi et al., 2024; Androshchuk, 2023).

Thus, the developed methods provide an integrated approach to multimodal analysis, considering all key criteria, which allows for increased analysis efficiency and improved user experience in various industries. They open new opportunities for creating more reliable and efficient systems that meet modern requirements for processing and analysing multimodal data.

5. Discussion

By presenting a set of techniques that not only combine and synchronise various data modalities but also maximise processing efficiency, customise user engagement, and protect data privacy, this study makes a significant contribution to the field of multimodal content analysis. The suggested paradigm exhibits thorough coverage of technological, cognitive, and regulatory aspects of user experience in interactive systems, in contrast to previous approaches that concentrated on discrete components or limited modality integration.

Significant improvements in analytical accuracy and system responsiveness are achieved by the synergistic integration of modalities through joint embedding and temporal-spatial synchronisation. This is the main innovation. This study's paradigm expands the applicability beyond edge computing to user-facing interactive systems with real-time requirements, in contrast to earlier models like those covered by Wang et al. (2022), who concentrated on bidirectional caching for multimodal communication in network contexts. The results demonstrate actual scalability for embedded systems, which has not been systematically addressed in similar studies, with a 37% reduction in memory consumption and a 10% reduction in energy usage.

Furthermore, the suggested optimisation methods – such as parallel processing, model pruning, and algorithmic quantization – outperform the computational efficiency standards frequently mentioned in the literature. Although interface design and emotional engagement were

highlighted in earlier research by Awada et al. (2018), the problem of resource limitations was not adequately addressed. Our research closes this gap by providing techniques that are both computationally efficient and cognitively responsive, which makes them appropriate for use in situations with limitations like wearable and mobile devices.

This study showed that new methods of multimodal content analysis can improve the user experience by increasing attention and efficiency of information delivery, while the work of Breitschaft et al. (2021) indicated that a haptic display with electrostatic plaster in an automotive environment provides an improved user experience, where multimodal feedback was preferred by most participants. The results of this study also demonstrated that multimodal content analysis can significantly increase the reliability of information and improve the overall user experience by integrating different modalities.

At the same time, the work of Ince et al. (2024) showed that the critical aspects of multimodal experience in autonomous vehicles are attention and duration of interaction. In addition, the results of the study showed that the provided methods of multimodal content analysis provide more accurate and faster data processing, which leads to improved user experience. In turn, Vijayakumar (2024) pointed out that emotionally aware interfaces improve user experience by integrating audio, haptic, and visual feedback, which also demonstrates the importance of a multimodal approach to improving the effectiveness of user interaction. This work highlighted that multimodal content analysis significantly improves user experience by better integrating different types of data, such as text, audio, and video, which provides a more complete and convenient interaction. Similarly, a study by Choudhary et al. (2024) showed that multimodal assessment of user-generated content, which considers text, audio, video, and images, is more effective in detecting fake news, suspicious profiles, and fake reviews, which also highlights the importance of using a multimodal approach to improve interaction and content analysis.

This work revealed that the integration of different modalities to optimize the user experience is key, like the work of Panwar (2024), who emphasized the importance of progressive applications in web development, in particular their impact on improving efficiency and user engagement. In this study, special attention was paid to analysing the results in various aspects, while Moradi's (2022) study highlighted the importance of understanding the socio-emotional dimension of active ageing users' interaction with information technology.

While this work found that multimodal warnings improve user trust and satisfaction, the study by Zhang and Tan (2021) found that multimodal warning types can also lead to higher levels of annoyance, especially when haptic cues are included, indicating their potential to improve driver performance in warning situations. Although this study demonstrated the importance of integrating multimodal interfaces to improve the user experience in interacting with interactive systems, the work of Santhosh et al. (2024) focused on a specific domain and found that sophisticated multimodal product rankings contribute to a high level of user trust and satisfaction, resulting in a more satisfying and rich online shopping or content consumption experience.

Moreover, the results of this study indicated the high efficiency of new methods for analysing multimodal content, which contributes to improving the user experience of interacting with multimedia systems. Similarly, a study by Manmothe & Jadhav (2024) confirmed that the integration of different modalities in multimedia content is important for improving sentiment analysis and providing effective cross-search, which is relevant in today's digital environment. Compared to the results of the study, which indicated the relevance of integrating different modalities to improve the user experience, the work of Pradipta et al. (2024) emphasized the importance of the developed application for people who love cinema, but noted various problems associated with its use and overall user satisfaction.

The methods of this work demonstrated increased user attention and optimized information delivery, which is comparable to the results of Xu et al. (2023), where real-time emotion monitoring also showed a significant improvement in user engagement and overall satisfaction. In contrast to this study, which indicated the possibility of practical use of methods with increased accuracy, speed, and confidentiality, which emphasizes their effectiveness in improving user experience, the results of Saputra et al. (2024) focused on a specific application with a high level of user satisfaction, in particular in terms of attractiveness, clarity, efficiency, reliability, stimulation, and novelty, which reflects the importance of high-quality interaction with the application to achieve significant user satisfaction.

While this work pointed to significant improvements in user experience using the provided content analysis techniques in multimodal interactive systems, the results of the study by Richas & Kamal (2024) confirm that mobile banking applications have a high level of user satisfaction in important aspects, which underlines their success in responding to the needs in this area. Compared to the present study, which focused on the integration of various multimodal media to enhance the user experience in technology-enabled environments, the study by Smith-Harvey & Aguayo (2024) confirmed that this approach promotes deeper engagement, critical thinking, and a more comprehensive understanding of certain complex concepts, which underlines its effectiveness in improving the learning process. In addition, this study found that multimodal data has limitations in size and sound, which makes it difficult to process and model it to improve the user experience in interactive systems. Mullick et al. (2024) also pointed out these limitations in long-term, multimodal data from mobile sensors. While this work demonstrated the results of optimizing multimodal data analysis in the context of interactive systems, the work of Ahmad (2024) considered the implementation of supported artificial intelligence solutions to improve the user experience by optimizing usability, automating tasks, and providing personalized experiences, which contributes to the efficiency and productivity of enterprises.

Finally, in comparison to this study, which confirmed significant progress in improving the user experience through the introduction of new methods of multimodal content analysis, the study by Kamil (2023) highlighted a similar moderate level of user satisfaction, but noted that there is a need to improve the efficiency and overall experience of using the LinkedIn platform. In addition, the results in this paper focus on the development of multimodal content analysis methods for optimizing user experience in interactive systems, aimed at integrating different types of data to improve their accuracy and efficiency, while Salim (2023) provides a comprehensive analysis of interaction with smartwatches, including user experience elements, feature usage, and emotional aspects, emphasizing the importance of integrating functional and emotional aspects to increase user satisfaction and adaptation to wrist devices.

Thus, methods of multimodal content analysis have proven to be effective in improving the user experience in interactive systems, especially due to the increased accuracy and efficiency of data processing. The results of the study confirm that the integration of different modalities contributes to improving the quality of user interaction and may be of considerable interest for further research and development in this area.

6. Conclusion

This study developed and validated a set of methods for multimodal content analysis aimed at enhancing the user experience in interactive systems. The conducted research demonstrated that the integration and synchronization of heterogeneous data modalities, including text, audio, video, and images, markedly improves the efficiency, accuracy, and robustness of data processing, ultimately resulting in higher levels of user satisfaction and engagement. Through both theoretical modeling and practical implementation across diverse application domains, the study established that multimodal analysis allows for a more comprehensive interpretation of user inputs, thereby offering greater personalization and responsiveness of digital systems.

Quantitative results confirmed that the proposed methods contribute to measurable improvements in user attention, reaction time, and perceived usability. These metrics are critical in assessing the effectiveness of interactive platforms. The developed approaches, such as optimized algorithmic pipelines, noise-resistant models, and adaptive personalization schemes, outperformed conventional unimodal techniques in terms of processing speed, resilience to incomplete data, and scalability. Furthermore, the integration of privacy-preserving mechanisms, including encryption, anonymization, and differential privacy techniques, supports regulatory compliance and strengthens user trust.

The study contributes to the literature by presenting a cohesive framework for multimodal content analysis that aligns advanced machine learning strategies with practical user-centered design. The originality of the work lies in the holistic integration of modalities via joint embedding mechanisms, the simultaneous optimization for computational efficiency, and the application of robust learning techniques to mitigate the effects of noisy or missing inputs. These contributions are particularly relevant in domains such as healthcare, education, streaming services, and e-commerce, where the quality of the user interface significantly impacts user outcomes and satisfaction.

However, several limitations were identified. First, the current implementation relies on pre-defined modality structures and may exhibit reduced performance when modalities are asynchronous or imbalanced in quality. Second, while the study addressed personalization, it did not fully explore the ethical implications of adaptive systems, such as potential algorithmic bias. Third, real-time processing capabilities were tested under laboratory conditions. Further validation in large-scale operational environments remains necessary to establish broader generalizability.

Future research should prioritize the development of advanced fusion models capable of handling modality-specific noise and asynchronous inputs. It is also important to enhance real-time processing capacities through hardware-accelerated frameworks or edge computing paradigms. Furthermore, greater attention should be paid to user trust, ethical transparency, and algorithmic fairness in personalized multimodal systems. Finally, there is a need to expand application testing in high-stakes environments such as telemedicine, autonomous vehicles, and financial decision-support systems.

Acknowledgement

None.

References

[1] Ahmad N, 2024. Revolutionizing ERP integration: AI-powered solutions for effortless usability and enhanced user experience. http://dx.doi.org/10.13140/RG.2.2.33419.30247

[2] Alzubi TM, Alzubi JA, Singh A & Alzubi OA 2023. A multimodal human-computer interaction for smart learning system. International Journal of Human-Computer Interaction. https://doi.org/10.1080/10447318.2023.2206758

[3] Androshchuk A 2023. Cyber protection in programming languages Java and C#. Technologies and Engineering, 24(3):9–14. https://doi.org/10.30857/2786-5371.2023.3.1

[4] Aviv I, Gafni R, Sherman S, Aviv B, Sterkin A & Bega E 2023. Cloud infrastructure from python code–breaking the barriers of cloud deployment. In: European Conference on Software Architecture, ECSA (pp.1–8). Available at: https://www.researchgate.net/profile/Itzhak-Aviv/publication/373897534_Cloud_Infrastructure_from_Python_Code_-breaking_the_Barriers_of_Cloud_Deployment/links/6501edd2808f9268d573dea5/Cloud-Infrastructure-from-Python-Code-breaking-the-Barriers-of-Cloud-Deployment.pdf

[5] Awada IA, Mocanu I & Florea AM 2018. Exploiting multimodal interfaces in eLearning systems. International Scientific Conference: eLearning and Software for Education Conference, 2:174–181. https://doi.org/10.12753/2066-026x-21-094

[6] Azieva G, Kerimkhulle S, Turusbekova U, Alimagambetova A & Niyazbekova S 2021. Analysis of access to the electricity transmission network using information technologies in some countries. E3S Web of Conferences, 258:11003. https://doi.org/10.1051/e3sconf/202125811003

[7] Bezshyyko O, Dolinskii A, Bezshyyko K, Kadenko I, Yermolenko R & Ziemann V 2008. PETAG01: A program for the direct simulation of a pellet target. Computer Physics Communications, 178(2):144–155. https://doi.org/10.1016/j.cpc.2007.07.013

[8] Bisenovna KA, Ashatuly SA, Beibutovna LZ, Yesilbayuly KS, Zagievna AA, Galymbekovna MZ & Oralkhanuly OB 2024. Improving the efficiency of food supplies for a trading company based on an artificial neural network. International Journal of Electrical and Computer Engineering, 14(4):4407–4417. https://doi.org/10.11591/ijece.v14i4.pp4407-4417

[9] Bouchey B, Castek J & Thygeson J 2021. Multimodal learning. In: Learning Environments in STEM Higher Education (pp.35–54). Cham: Springer. https://doi.org/10.1007/978-3-030-58948-6_3

[10] Breitschaft SJ, Pastukhov A & Carbon C-C 2021. Where's my button? Evaluating the user experience of surface haptics in featureless automotive user interfaces. IEEE Transactions on Haptics, 15(2):292–303. https://doi.org/10.1109/TOH.2021.3131058

[11] Chaiyaphum T 2022. Developing supply chain innovation model for operation and adaptation of logistics service providers in the disruptive era. Available at: http://www.dspace.spu.ac.th/handle/123456789/8697

[12] Chiu WHK, Ko WSK, Cho WCS, Hui SYJ, Chan WCL & Kuo MD 2024. Evaluating the diagnostic performance of large language models on complex multimodal medical cases. Journal of Medical Internet Research, 26:e53724. https://doi.org/10.2196/53724

[13] Choudhary M, Singh S & Rathore SS 2024. Beyond text: Multimodal credibility assessment approaches for online user-generated content. ACM Transactions on Intelligent Systems and Technology. https://doi.org/10.1145/3673236

[14] Dewangan O & Mishra M 2023. TLMAEMS: Design of an efficient transfer learning model with auto encoders for multimodal sentiment analysis via deep sentiment networks. European Chemical Bulletin, 12(Special Issue 5):1453–1467. http://dx.doi.org/10.48047/ecb/2023.12.si5.1362023.30/05/2023

[15] Fang H, Liang J & Sha L 2024. Enhanced multimodal recommendation systems through reviews integration. Research Square. https://doi.org/10.21203/rs.3.rs-4333408/v1

[16] Henrique BM, Sobreiro VA & Kimura H 2023. Practical machine learning: Forecasting daily financial markets directions. Expert Systems with Applications, 233:120840. https://doi.org/10.1016/j.eswa.2023.120840

[17] Heraz A, Chen L & Bhyravabhottla KKA 2024. Enhancing user receptivity in e-learning environments through gesture-based emotion tracking as a pathway to mental well-being and health equity. JMIR Formative Research. https://doi.org/10.2196/preprints.57293

[18] Iklassova K, Aitymova A, Kopnova O, Shaporeva A, Abildinova G, Nurbekova Z, Almagambetova L, Gorokhov A & Aitymov Z 2024. Ontology modeling for the automation of questionnaire data processing. Eastern-European Journal of Enterprise Technologies, 5(2-131):36–52. https://doi.org/10.15587/1729-4061.2024.314129

[19] Ince EB, Cha K & Cho J 2024. Towards a conceptual model of users' expectations of an autonomous in-vehicle multimodal experience. Human Behavior and Emerging Technologies, 2024(1):7418597. https://doi.org/10.1155/2024/7418597

[20] Jagnade G, Sable S & Ikar M 2023. Advancing multimodal fusion in human-computer interaction: Integrating eye tracking, lips detection, speech recognition, and voice synthesis for intelligent cursor control and auditory feedback. International Conference on Computing Communication and Networking Technologies. https://doi.org/10.1109/ICCCNT56998.2023.10306457

[21] Kamil MAF 2023. User Experience analysis of linkedin social media using usability metric for user experience (UMUX). Journal of Information Engineering and Educational Technology, 7(2):78–82. https://doi.org/10.26740/jieet.v7n2.p78-82

[22] Kamtab P & Wiriyanon T 2022. Design a digital multimodal user interface for senior citizens. Journal of Education Studies, 50(1):6. https://doi.org/10.58837/chula.educu.50.1.5

[23] Kerimkhulle S & Aitkozha Z 2017. A criterion for correct solvability of a first order difference equation. AIP Conference Proceedings, 1880:040016. https://doi.org/10.1063/1.5000632

[24] Kerimkhulle S, Kerimkulov Z, Bakhtiyarov D, Turtayeva N & Kim J 2021. In-Field Crop-Weed Classification Using Remote Sensing and Neural Network. In: SIST 2021 - 2021 IEEE International Conference on Smart Information Systems and Technologies (9465970). https://doi.org/10.1109/SIST50301.2021.9465970.

[25] Kerimkhulle S, Obrosova N, Shananin A & Tokhmetov A 2023. Young Duality for Variational Inequalities and Nonparametric Method of Demand Analysis in Input–Output Models with Inputs Substitution: Application for Kazakhstan Economy. Mathematics, 11(19):4216. https://doi.org/10.3390/math11194216

[26] Khoda V, Leshchuk N, Topalov A, Robotko S, Klymenko O & Nekrasov S 2024. Computerized Lathe Control System based on Internet of Things Technology. In: Proceedings - International Conference on Advanced Computer Information Technologies, ACIT (pp.674–677). Ceske Budejovice: Institute of Electrical and Electronics Engineers. https://doi.org/10.1109/ACIT62333.2024.10712548

[27] Khrulov M 2025. Analysis of the current state of computer systems in the field of virtual healthcare. Technologies and Engineering, 26(1): 55–66. https://doi.org/10.30857/2786-5371.2025.1.5

[28] Kolodiziev O, Krupka M, Shulga N, Kulchytskyy M & Lozynska O 2021. The level of digital transformation affecting the competitiveness of banks. Banks and Bank Systems, 16(1):81–91. https://doi.org/10.21511/bbs.16(1).2021.08

[29] Konurbayeva ZT, Madiyarova ES & Rakhimberdinova MU 2015. Algorithm for generating competitive potential of engineering within the regional economy. Actual Problems of Economics, 168(6):236–247.

[30] Kovalchuk D 2025. Utilising large language models for automated real-time cyber threat analysis. Bulletin of Cherkasy State Technological University, 30(1):48–58. https://doi.org/10.62660/bcstu/1.2025.48

[31] Maksymov A & Tryus Yu 2025. Information technology for solving the multi-criteria decision-making problem using the modified Fuzzy TOPSIS method. Bulletin of Cherkasy State Technological University, 30(1):91–106. https://doi.org/10.62660/bcstu/1.2025.91

[32] Manmothe SR & Jadhav JR 2024. Integrating multimodal data for enhanced analysis and understanding: Techniques for sentiment analysis and cross-modal retrieval. Journal Of Advanced Zoology, 45(S4):22–28. https://doi.org/10.53555/jaz.v45iS4.4144

[33] Mao X & Zhang J 2021. On the power of SVD in the stochastic block model. https://doi.org/10.48550/arXiv.2309.15322

[34] Marques J & Marques RP 2023. Digital transformation of the hotel industry. Cham: Springer.

[35] Meier S 2022. Digital storytelling: A didactic approach to multimodal coherence. Frontiers in Communication, 7:906268. https://doi.org/10.3389/fcomm.2022.906268

[36] Moradi P 2022. Ageing with technology – An extended multimodal design study of active ageing users' emotional experience with social robots. Available at: https://hdl.handle.net/10292/15512

[37] Mullick T, Shaaban S, Radovic A & Doryab A 2024. Framework for ranking machine learning predictions of limited, multimodal, and longitudinal behavioral passive sensing data: Combining user-agnostic and personalized modeling. JMIR AI, 3:e47805. https://doi.org/10.2196/47805

[38] Naegelin M, Weibel RP, Kerr JI, Schinazi VR, La Marca R, Von Wangenheim F, Hoelscher C & Ferrario A 2023. An interpretable machine learning approach to multimodal stress detection in a simulated office environment. Journal of Biomedical Informatics, 139:104299. https://doi.org/10.1016/j.jbi.2023.104299

[39] Orazbayev B, Zhumadillayeva A, Kabibullin M, Crabbe MJC, Orazbayeva K & Yue X 2023. A Systematic Approach to the Model Development of Reactors and Reforming Furnaces With Fuzziness and Optimization of Operating Modes. IEEE Access, 11:74980–74996. https://doi.org/10.1109/ACCESS.2023.3294701

[40] Panwar V 2024. Leveraging progressive web apps (PWAs) for enhanced user experience and performance: A comprehensive analysis. International Journal of Management IT and Engineering, 14(4):31–43.

[41] Pavlova D, Dovramadjiev T, Daskalov D, Mirchev N, Peev I, Radeva J, Dimova R, Kavaldzhieva K, Mrugalska B, Szabo G, Kandioglou A 2024. 3D Design of a Dental Crown with Artificial Intelligence Based in Cloud Space. Lecture Notes in Networks and Systems, 817:437–445. https://doi.org/10.1007/978-981-99-7886-1_37

[42] Porkodi S & Raman AM 2025. Success of cloud computing adoption over an era in human resource management systems: a comprehensive meta-analytic literature review. Management Review Quarterly, 75(2):1041–1075. https://doi.org/10.1007/s11301-023-00401-0

[43] Pradipta RAH, Widiani HN, Tama DA, Ramdani AL, Komariansyah KK, Nurfajri MI, Barus I & Fami A 2024. Analyzing user experience in the M-TIX application using the user experience questionnaire method. International Journal of Multilingual Education and Applied Linguistics, 1(3):61–71. https://doi.org/10.61132/ijmeal.v1i3.62

[44] Ragan LC & Villarin LJR 2021. Emergent guiding principles for STEM education. In: Innovative Learning Environments in STEM Higher Education (pp.107–119). Cham: Springer. https://doi.org/10.1007/978-3-030-58948-6_6

[45] Ramtohul A & Khedo KK 2023. An engaging user experience framework for mobile augmented reality. Research Square. https://doi.org/10.21203/rs.3.rs-3080680/v1

[46] Rexhepi BR, Kumar A, Gowtham MS, Rajalakshmi R, Paikaray MD & Adhikari PK 2023. An Secured Intrusion Detection System Integrated with the Conditional Random Field For the Manet Network. International Journal of Intelligent Systems and Applications in Engineering, 11(3s):14–21.

[47] Richas F & Kamal I 2024. Comparison analysis of user experience on m-BCA and BRImo mobile banking applications using the user experience questionnaire (UEQ) method. International Journal of Management and Business Economics, 2(3):47–51. https://doi.org/10.58540/ijmebe.v2i3.551

[48] Rust RT 2020. The future of marketing. International Journal of Research in Marketing, 37(1):15–26. https://doi.org/10.1016/j.ijresmar.2019.08.002

[49] Salim F 2023. Identifying user experience (UX) elements and emotional experiences on smartwatch devices. http://dx.doi.org/10.22677/THESIS.200000726145

[50] Santhosh P, Ashweej A, Agarwal G & Raj GD 2024. Personalized product ranking system for enhanced user experience. MATEC Web of Conferences, 392:01077. https://doi.org/10.1051/matecconf/202439201077

[51] Saputra FA, Yeza MP, Fahrezy MF, Fami A & Barus IRG 2024. Article comparing user experience maxim and in drive applications using the user experience questionnaire. ILKOMNIKA Journal of Computer Science and Applied Informatics, 6(1):79–90. https://doi.org/10.28926/ilkomnika.v6i1.614

[52] Smith-Harvey J & Aguayo C 2024. Modes of meaning: Multimodal media & 4E+ cognition in tech-enhanced learning. Pacific Journal of Technology Enhanced Learning, 6(1):10–11. https://doi.org/10.24135/pjtel.v6i1.181

[53] Taylor DL, Yeung M & Bashet AZ 2021. Personalized and adaptive learning. In: Innovative Learning Environments in STEM Higher Education (pp.17–34). Cham: Springer. https://doi.org/10.1007/978-3-030-58948-6_2

[54] Tkachenko O, Chechet A, Chernykh M, Bunas S & Jatkiewicz P 2025. Scalable Front-End Architecture: Building for Growth and Sustainability. Informatica (Slovenia), 49(1):137–150. https://doi.org/10.31449/inf.v49i1.6304

[55] Tokarieva KS, Kovalchuk OY, Kolesnikov AP, Dzyurbel AD, Bodnar-Petrovska OB & Predmestnikov OG 2024. The use of ai-language models in judicial proceedings: information and legal aspects. Revista Juridica, 2(78):520–538. https://doi.org/10.26668/revistajur.2316-753X.v2i78.6928

[56] UI/UX staffing 2019. Available at: https://evergreens.com.ua/ua/articles/ui-and-ux-personalization.html

[57] Varanitskyi D, Rozkolodko O, Liuta M, Zakharova M & Hotunov V 2024. Analysis of data protection mechanisms in cloud environments. Technologies and Engineering, 25(1):9–16. https://doi.org/10.30857/2786-5371.2024.1.1

[58] Vijayakumar T 2024. Enhancing user experience through emotion-aware interfaces: A multimodal approach. Journal of Innovative Image Processing, 6(1):27–39. https://doi.org/10.36548/jiip.2024.1.003

[59] Wang C, Yu X, Xu L, Wang Z & Wang W 2022. Multimodal semantic communication accelerated bidirectional caching for 6G MEC. Future Generation Computer Systems, 140:225–237. https://doi.org/10.1016/j.future.2022.10.036
[60] What is Data Protection and Privacy? 2024. Available at: https://cloudian.com/guides/data-protection/data-protection-and-privacy-7-ways-to-protect-user-data/#:~:text=By%20protecting%20data%2C%20companies%20can,%2C%20encryption%2C%20and%20endpoint%20protection
[61] Wildfeuer J & Lehmann C 2024. Drawing multimodality's bigger picture: Metalanguages and corpora for multimodal analyses. Lausanne: Frontiers Media SA. https://doi.org/10.3389/978-2-8325-5196-7
[62] Xhafka E, Sinoimeri D & Teta J 2024. Evaluating the Impact of E-Governance on Public Service Improvement in Albania: A Quantitative Analysis. Sustainability (Switzerland), 16(24):10896. https://doi.org/10.3390/su162410896
[63] Xu Y, Lin Y-S, Zhou X & Shan X 2023. Utilizing emotion recognition technology to enhance user experience in real-time. Computing and Artificial Intelligence, 2(1):1388. https://doi.org/10.59400/cai.v2i1.1388
[64] Yasinthorn Y & Celadon S 2023. Security enhancement and performance optimization for WordPress-based Websites, a case study of WorayuthIT website. Available at: http://digital.csmsu.net:8080/library/handle/123456789/218
[65] Yasmin S & Jamal CQ 2023. Multisensory modeling of tabular data for enhanced perception and immersive experience. In: Advances in Visual Computing (pp.452–465). Cham: Springer. https://doi.org/10.1007/978-3-031-47966-3_36
[66] Zhang Y & Tan H 2021. Effects of multimodal warning types on driver's task performance, physiological data and user experience. In: Cross-Cultural Design (pp.304–315). Cham: Springer. https://doi.org/10.1007/978-3-030-77080-8_25
[67] Zhetenbayev N, Zhauyt A, Balbayev G & Shingissov B 2022. Robot device for ankle joint rehabilitation: A review. Vibroengineering Procedia, 41:96–102. https://doi.org/10.21595/vp.2022.22507
[68] Ziker C, Truman B & Dodds H 2021. Cross Reality (XR): Challenges and opportunities across the spectrum. In: Innovative Learning Environments in STEM Higher Education (pp.55–77). Cham: Springer. https://doi.org/10.1007/978-3-030-58948-6_4