ISSN: 3105-7861
E-ISSN: 3105-787X

UNEC JOURNAL OF Computer Science and Digital Technologies

JCSDT

Volume 2, № 1, June 2026
journals.unec.edu.az

UNEC JOURNAL OF
Computer Science and Digital Technologies

JCSDT

Volume 2, № 1, June 2026

Baku - Azerbaijan

Publishing office: 6 Istiglaliyyat str., Baku AZ1001, Azerbaijan

Azerbaijan State University of Economics (UNEC)

Phone: (+994 12) 437 10 86

(+994 12) 492 65 09

Fax: (+994 12) 492 59 40

Website: www.unec.edu.az

email: unec@unec.edu.az

Number of copies: 100 pieces

The journal has been published by the "İqtisad Universiteti" publishing house of UNEC

Official website of the journal: https://journals.unec.edu.az/jcsdt

Foreword

Dear Colleagues and Members of the Academic Community,

On behalf of the Editorial Board, we are pleased to present the second issue of the *Journal of Computer Science and Digital Technologies (JCSDT)*, published by the Department of Digital Technologies and Applied Informatics at Azerbaijan State University of Economics (UNEC). Following the positive academic resonance of the inaugural issue, this volume reflects the journal's growing maturity, the expansion of its scholarly scope, and the strengthening of its position within the international academic discourse.

JCSDT operates in close alignment with Azerbaijan's national digital development priorities, particularly the **State Program for the Development of Artificial Intelligence (2025–2028)** and the **Digital Development Concept**. In line with these strategic frameworks, the journal seeks to promote evidence-based research in artificial intelligence, data science, cybersecurity, software engineering, human–computer interaction, and other advanced digital technologies that contribute to sustainable innovation and digital transformation.

JCSDT is a biannual, English-language, open-access journal that adheres to rigorous peer-review and editorial standards. This issue features contributions from experienced UNEC scholars as well as researchers from the international academic community. The journal's primary readership includes university faculty, researchers, postgraduate students, and professionals from both public and private sectors who are actively engaged in digital transformation processes.

Looking ahead, JCSDT prioritizes the continuous improvement of its editorial and peer-review processes, indexing in reputable international academic databases, the publication of special issues on emerging topics, and strengthened synergy with international conferences. Through its open-access publishing model under a Creative Commons license, the journal aims to ensure broad accessibility of scholarly knowledge and to enhance its practical, societal, and scientific impact.

We extend our sincere gratitude to all authors, reviewers, and members of the Editorial Board whose dedication and professionalism have contributed to the realization of this issue. We are confident that the *Journal of Computer Science and Digital Technologies (JCSDT)* will continue to serve as a reputable academic platform that fosters innovation, strengthens scholarly collaboration, and bridges regional initiatives with the global digital research agenda.

Yours sincerely,
The Editorial Board

UNEC Journal of Computer Science and Digital Technologies
Volume 2, № 1, 2026, pp. 5-15
https://doi.org/10.30546/UNECCSDT.2026.001.210

Section:
Artificial Intelligence &
Machine Learning

Efficient and Accurate Potato Disease Classification Using Lightweight Vision Transformers: A Comparative Benchmark Against a Deep CNN Architecture

Javanshir Zeynalov[1]*, Yiğitcan Çakmak[2]
[1]Faculty of Architecture and Engineering, Nakhchivan State University, Nakhchivan, Azerbaijan*
[2]Faculty of Engineering, Igdir University, Igdir, Türkiye
[1]0009-0002-4985-6371, cavansirzeynalov@ndu.edu.az
[2]0009-0008-7227-9182, ygtcncakmak@gmail.com

Abstract

Potato is among the most important staple crops underpinning global food security; however, its productivity remains highly vulnerable to destructive foliar diseases, particularly Early Blight and Late Blight, which can cause substantial yield losses when not detected and managed in a timely manner. Conventional disease diagnosis largely depends on visual inspection by farmers or specialists, a process that is time-consuming, expertise-dependent, and prone to subjective interpretation, especially under field conditions where symptoms may overlap or appear at early infection stages. To address these limitations, this study proposes and evaluates an automated deep learning-based framework for classifying potato leaf images into three categories: healthy, Early Blight, and Late Blight, using the publicly available PlantVillage dataset. A comparative assessment was conducted between a well-established convolutional architecture, ResNet-101, and two Vision Transformer-based models, namely Swin Base and MobileViT v2. The models were evaluated in terms of classification effectiveness using accuracy, precision, recall, and F1-score, while their computational practicality was examined through parameter count and GFLOPs. The experimental findings indicate that although all architectures achieved strong diagnostic performance, the transformer-based models consistently surpassed the conventional CNN baseline. Among them, MobileViT v2 delivered the best overall performance, reaching a test accuracy of 99.69% while maintaining a highly compact architecture with only 4.39 million parameters. This combination of high predictive accuracy and low computational demand suggests that lightweight Vision Transformer models offer a more practical and efficient alternative to deeper CNN-based approaches for potato disease recognition. These results underline the potential of such architectures for deployment in mobile, embedded, or other resource-constrained agricultural diagnostic systems, supporting more timely disease management and contributing to sustainable precision farming practices.

Keywords: Potato diseases, Deep learning, Vision Transformer (ViT), Plant disease classification, Computational efficiency

Received:	*Revised:*	*Accepted:*	*Published:*
20/05/2026	*03/06/2026*	*06/06/2026*	*17/06/2026*

1. Introduction

Potato (*Solanum tuberosum*) ranks among the most widely cultivated and consumed food crops worldwide, following maize, wheat, and rice in global importance, and it occupies a central position in sustaining the international food supply chain [1]. Owing to its high productivity, broad adaptability, and substantial nutritional value, potatoes serve as a major staple for more than one billion people and contribute directly to global food security [2]. Despite this importance, potato production remains highly susceptible to a range of destructive diseases, many of which are caused by fungal and bacterial pathogens and can severely reduce both yield and crop quality [3]. Among these, Early blight (*Alternaria solani*) and Late blight (*Phytophthora infestans*) are particularly damaging, as uncontrolled infections may lead to extensive economic losses and, in severe cases, the failure of entire harvests [4].

In conventional agricultural practice, the detection and management of potato diseases have largely relied on visual inspection performed by farmers, agronomists, or plant protection specialists [5]. Although widely used, this approach is inherently labor-intensive, subjective, and vulnerable to diagnostic errors, particularly during the early stages of infection when visual symptoms are subtle, overlapping, or easily mistaken for abiotic stress and nutrient-related disorders [6]. As a result, diagnostic reliability often depends heavily on the experience of the evaluator, leading to variability across observers and field conditions. These limitations may delay timely intervention or cause inappropriate fungicide application, increasing production costs while also intensifying environmental risks associated with excessive or unnecessary chemical use [7]. Therefore, the development of automated, accurate, and scalable diagnostic systems has become an increasingly important requirement for effective potato disease management [8].

Recent advances at the intersection of computer vision and deep learning have introduced powerful data-driven approaches for addressing long-standing challenges in precision agriculture [9,10]. These developments have been particularly influential in automated plant disease recognition, where image-based deep learning models have shown strong potential for accurate and scalable crop health assessment [11–14]. In this context, Convolutional Neural Networks (CNNs) have been widely adopted as a dominant methodological framework due to their ability to learn hierarchical and discriminative visual representations directly from image data [15,16]. Their strong performance across diverse agricultural imaging tasks has established CNNs as a reliable baseline for plant disease classification [17–19]. More recently, Vision Transformers (ViTs) have emerged as a compelling alternative architecture by replacing purely convolutional feature extraction with self-attention mechanisms capable of modeling long-range dependencies across image regions. Following their success in general-purpose computer vision, ViT-based models are increasingly being explored for domain-specific applications, including agricultural image analysis and crop disease diagnosis [20].

Within this context, the present study develops and evaluates a deep learning-based framework for the automated classification of three potato leaf conditions—healthy, Early blight, and Late blight—using the publicly available PlantVillage dataset. A systematic comparative analysis is conducted between a representative deep CNN architecture, ResNet-101, and two transformer-based models, Swin Base and MobileViT v2, to assess both predictive performance and practical computational suitability. Model effectiveness is evaluated using standard classification metrics, including accuracy, precision, recall, and F1-score, while computational efficiency is considered through model complexity indicators. By comparing established convolutional learning with more recent transformer-based paradigms, this study aims to provide evidence on the suitability of lightweight and high-performing architectures for practical potato disease diagnosis. The resulting findings are expected to support the development of accessible, efficient, and scalable decision-support tools for precision crop protection, thereby contributing to more sustainable potato production and improved agricultural resilience.

2. Related Work

Wang and Su (2024) [21] provided a broad review of deep learning applications across the potato production chain, organizing existing studies around major operational domains such as crop health monitoring, yield estimation, and resource management. Their review covered a range of architectures, including Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), and discussed their use in tasks extending from pest and disease detection to market-oriented applications such as price forecasting. Although the study emphasized the considerable potential of deep learning to improve productivity, decision-making, and operational efficiency in potato production, it also highlighted persistent barriers to real-world adoption, particularly the limited availability of diverse datasets and the difficulty of deploying these systems under practical field conditions.

Selvi et al. (2024) [22] introduced CropViT, a lightweight Vision Transformer architecture designed for efficient crop disease diagnosis. The model was fine-tuned and evaluated on nine crop categories from the PlantVillage dataset and was directly compared with a conventional CNN baseline. Their results showed that CropViT achieved a mean accuracy of 98.64%, outperforming the CNN model and demonstrating the capacity of transformer-based architectures to deliver high diagnostic performance while maintaining computational efficiency. This work provided important evidence that compact transformer models can serve as practical alternatives to traditional convolutional approaches in agricultural disease recognition.

Dutta et al. (2024) [23] developed a customized CNN model for the early detection and classification of potato blight diseases using leaf images. Their study focused on distinguishing healthy leaves from Early blight and Late blight cases and compared the proposed architecture with established deep learning models, including ResNet-50, VGG-16, and GoogLeNet. The customized CNN achieved an accuracy of 98%, surpassing the benchmark models and indicating that task-specific architectural design can improve performance in potato disease classification. Their findings further support the relevance of domain-adapted deep learning models for agricultural phytopathology applications.

Bajpai et al. (2024) [24] improved the standard Swin Transformer architecture for potato leaf disease detection by incorporating a customized sequential classification head consisting of Linear, ReLU, and Dropout layers. This modification was intended to strengthen feature representation and reduce overfitting during classification. Evaluated on a custom potato leaf dataset including Early blight and Late blight categories, the enhanced Swin Transformer reached an accuracy of 99.38%. The study demonstrated that targeted modifications to transformer-based models can improve generalization and classification accuracy in specialized agricultural computer vision tasks.

Zhang et al. (2025) [25] addressed both model efficiency and diagnostic accuracy in potato leaf disease recognition by proposing VGG16S, an optimized variant of the original VGG16 architecture. Their approach combined several architectural refinements, including the replacement of fully connected layers with global average pooling, the integration of the CBAM attention mechanism, and the adoption of the Leaky ReLU activation function. These changes substantially reduced the number of parameters to approximately one-tenth of the original VGG16 model while increasing classification accuracy to 97.87% on their proprietary dataset. Their findings illustrate the value of architectural compression and attention-based enhancement for developing more efficient plant disease diagnosis systems.

Sharma and Sharma (2024) [26] investigated the use of Recurrent Neural Networks for classifying healthy and diseased potato leaves from RGB images in the PlantVillage dataset. Their proposed model employed Long Short-Term Memory units for feature extraction and was compared with both a CNN and a Feedforward Neural Network. The RNN-based architecture achieved an accuracy of 92.7%, outperforming the alternative models evaluated in the study. Although convolutional architectures remain more common in image-based plant disease classification, their results suggest that sequence-oriented models may still offer useful representational capabilities when adapted to visual classification tasks.

Zoralioğlu and Polat (2024) [27] examined the influence of data augmentation and class balancing on potato disease classification performance. They evaluated a custom 5-layer CNN, EfficientNetB2, and ConvNeXtSmall on the PlantVillage dataset under both imbalanced and augmented balanced conditions. Their results revealed that model performance was strongly affected by the underlying data distribution. While the custom CNN performed best on the original imbalanced dataset, EfficientNetB2 achieved the highest accuracy of 99.89% after augmentation and class balancing. This finding highlights the importance of preprocessing and dataset balancing strategies for enabling advanced deep learning architectures to reach their full potential in plant disease detection.

Although previous studies have demonstrated the effectiveness of deep learning models for potato disease classification, several limitations remain evident in the existing literature. Many studies focus primarily on maximizing classification accuracy, while computational efficiency and deployment feasibility are often treated as secondary considerations. In addition, prior works frequently evaluate either CNN-based models or transformer-based architectures in isolation, limiting the ability to draw a balanced comparison between established convolutional methods and emerging attention-based approaches. Considering the practical requirements of precision agriculture, especially the need for accurate models that can operate on mobile or resource-constrained devices, there is still a need for comparative evaluations that jointly examine predictive performance and computational complexity. To address this gap, the present study systematically benchmarks ResNet-101, Swin Base, and MobileViT v2 on the PlantVillage potato leaf dataset, providing a performance–efficiency perspective for identifying architectures suitable for practical potato disease diagnosis.

3. Materials and Methods

3.1 Dataset and Data Preprocessing

This study employed the publicly available PlantVillage dataset, a widely used image repository for plant disease recognition tasks. From this dataset, the potato leaf subset was selected, comprising three diagnostic categories: healthy leaves, Early blight, and Late blight [28]. Representative samples from each class are presented in Figure 1, illustrating the visual characteristics associated with the corresponding leaf conditions and providing an overview of the classification targets addressed in this study. To enable a reliable and unbiased assessment of model performance, the dataset was divided into training, validation, and test subsets using a 70:15:15 ratio. The detailed distribution of images across these subsets is reported in Table 1, which summarizes the number of samples assigned to each class within the training, validation, and testing partitions.

Table 1. *Distribution of Classes Across Data Splits*

Class	Train (70%)	Validation (15%)	Test (15%)	Total
Early Blight	700	150	150	1000
Healthy	106	22	24	152
Late Blight	700	150	150	1000
Total	1506	322	324	2152

A consistent preprocessing pipeline was applied prior to model training to ensure input standardization across all architectures. Each image was resized to 224 × 224 pixels, matching the required input resolution of the pre-trained models used in this study. Pixel intensity values were subsequently scaled to the [0, 1] range, a normalization step intended to improve numerical stability and support more efficient convergence during training. To reduce the risk of overfitting and enhance the models' ability to generalize beyond the training data, data augmentation was applied only to the training subset.

Figure 1. Sample Images of Potato Leaves for Healthy, Early Blight, and Late Blight Classes

This augmentation process generated additional variability through random transformations, including horizontal flipping, rotation, and zooming, thereby exposing the models to a broader range of plausible visual variations while preserving the underlying class labels [29,30].

3.2 Foundational Principles of Deep Learning Models

3.2.1 Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (CNNs) are a class of deep learning architectures specifically designed to process structured grid-like data, with images being one of their most prominent application domains. Their effectiveness stems from a hierarchical feature learning mechanism, in which lower layers capture simple visual patterns such as edges and textures, while deeper layers progressively encode more complex and semantically meaningful representations. The core component of a CNN is the convolutional layer, where learnable kernels are applied across the input image to produce feature maps that highlight relevant spatial patterns. These convolutional operations are commonly followed by pooling layers, which reduce the spatial resolution of the feature maps and thereby lower computational cost while improving robustness to minor spatial variations. After successive stages of convolution and pooling, the extracted high-level features are passed to fully connected layers or classification heads, which transform the learned representations into final class predictions.

3.2.2 Vision Transformers (ViTs)

Vision Transformers (ViTs) depart from the conventional convolution-based design by adapting the Transformer architecture, originally developed for sequence modeling in natural language processing, to visual recognition tasks. Rather than relying on local receptive fields to process pixel neighborhoods, a ViT first divides an input image into a set of fixed-size, non-overlapping patches and treats these patches as a sequence of visual tokens. Each patch is flattened and projected into a latent embedding space, while learnable positional embeddings are added to preserve spatial information that would otherwise be weakened during tokenization. The resulting token sequence is then passed through multiple Transformer encoder layers, where multi-head self-attention allows the model to estimate relationships among all image patches simultaneously. This mechanism enables ViTs to capture long-range dependencies and global contextual patterns that may be difficult to model using purely local convolutional operations. The final representation is subsequently processed by a classification head, which maps the learned image-level features to the corresponding disease category.

3.3 Transfer Learning and Data Augmentation Strategy

Transfer learning was adopted to improve convergence efficiency and strengthen classification performance. The selected architectures were initialized with ImageNet pre-trained weights, enabling the models to benefit from generic visual representations learned from large-scale and diverse image collections. For each architecture, the original classification layer was removed and replaced with a new task-specific output layer configured for the three potato leaf categories considered in this study. The training procedure was implemented in two stages. In the first stage, the pre-trained feature extraction backbone was kept frozen, and only the newly added classification layer was trained to adapt the model to the target task. In the second stage, the full network was fine-tuned end-to-end using a low learning rate, allowing the pre-trained representations to be gradually adjusted to the visual characteristics of potato leaf diseases. This transfer learning strategy was complemented by data augmentation, which increased the diversity of the training samples and helped reduce overfitting, thereby improving the robustness and generalization capacity of the models.

3.4 Experimental Design and Training Protocol

A standardized experimental protocol was designed to ensure a fair and reproducible comparison among the selected deep learning architectures. All models were implemented in Python using the TensorFlow framework, and both training and inference procedures were carried out on a high-performance workstation equipped with an NVIDIA GeForce RTX 5090 GPU. To maintain consistency across experiments, the same training configuration was applied to all models. Model parameters were optimized using the Adam optimizer, with a learning rate of 1×10^{-4} during the fine-tuning stage. Training was performed with a mini-batch size of 16 for a maximum of 100 epochs. To limit overfitting and prevent unnecessary training, early stopping was applied with a patience value of 10 epochs, terminating the process when no improvement in validation loss was observed over consecutive epochs. The checkpoint corresponding to the lowest validation loss was retained and used for the final evaluation on the independent test set.

3.5 Performance Evaluation Metrics

The classification performance of each model was evaluated on the independent hold-out test set using a set of widely accepted evaluation metrics. Accuracy was used as the primary indicator of overall predictive performance, reflecting the proportion of correctly classified samples among all test instances. However, to obtain a more detailed understanding of model behavior, Precision, Recall, and F1-score were also computed. Precision measures the reliability of positive predictions by indicating the proportion of correctly identified positive samples among all samples predicted as positive. Recall, also referred to as sensitivity, quantifies the model's ability to detect all relevant positive instances within a given class. The F1-score combines Precision and Recall through their harmonic mean, providing a balanced performance measure that is particularly useful when class distributions are uneven. The mathematical formulations of these metrics are provided as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{2}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3}$$

$$\mathrm{F1-Score} = 2 * \frac{\mathrm{Precision} * \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}} \tag{4}$$

In these equations, TP, TN, FP, and FN represent the numbers of true positive, true negative, false positive, and false negative predictions, respectively. Since the present task involves a multi-class classification setting, the metrics were first calculated separately for each class and then aggregated using macro-averaging. This approach assigns equal importance to each class and produces a single overall score that reflects the model's classification performance across all potato leaf categories.

4. Results and Discussion

The experimental evaluation was conducted to systematically assess and compare the performance of the selected deep learning architectures in the task of potato leaf disease classification. Quantitative findings, including both classification of performance metrics and computational complexity indicators, are presented to provide a comprehensive basis for model comparison. Accuracy, precision, recall, and F1-score were calculated on the independent test set, thereby enabling an objective evaluation of each model's generalization capability on previously unseen data. The complete results are summarized in Table 2 and form the basis for the comparative analysis discussed in the following section.

Table 2. *Performance Evaluation Results of Deep Learning Models*

Models	Accuracy	Precision	Recall	F1-score	Params	GFLOPs
MobileViT v2 [31]	0.9969	0.9978	0.9861	0.9918	4.39M	2.8234
ResNet-101 [32]	0.9846	0.9772	0.9772	0.9772	42.51M	15.7288
Swin Base [33]	0.9938	0.9956	0.9839	0.9896	86.75M	30.3375

ResNet-101 was used as the representative CNN-based baseline in the comparative evaluation. The model achieved an accuracy of 0.9846, with precision, recall, and F1-score values of 0.9772, indicating a strong capacity to extract discriminative visual features from potato leaf images. This result confirms the effectiveness of deep residual learning for plant disease classification. Nevertheless, the model's computational cost remains relatively high, with 42.51 million parameters and 15.7288 GFLOPs, making it less suitable for resource-limited deployment scenarios when compared with more compact architectures. Although ResNet-101 provided a robust baseline, its performance was consistently exceeded by the transformer-based models, suggesting that attention-driven architectures may offer greater representational effectiveness for this classification task.

The Vision Transformer-based models achieved superior overall performance compared with the conventional CNN architecture. Swin Base reached an accuracy of 0.9938 and a precision of 0.9956, demonstrating the effectiveness of its hierarchical structure and shifted-window self-attention mechanism in capturing disease-relevant visual patterns. However, this high predictive performance was accompanied by the greatest computational demand among the evaluated models, requiring 86.75 million parameters and 30.3375 GFLOPs. By contrast, MobileViT v2 provided the most favorable balance between accuracy and efficiency. It achieved the highest accuracy of 0.9969, together with a precision of 0.9978 and an F1-score of 0.9918, while maintaining a substantially smaller computational footprint. The confusion matrix presented in Figure 2 further supports this result by showing that MobileViT v2 correctly classified all Early Blight and Late Blight samples, with 150 correct predictions in each category. Only one misclassification was observed in the Healthy class, where a single image was predicted as Late Blight, resulting in 23 correctly classified samples out of 24. This near-perfect classification pattern indicates that MobileViT v2 can effectively distinguish subtle visual differences among healthy and diseased potato leaves while preserving computational efficiency.

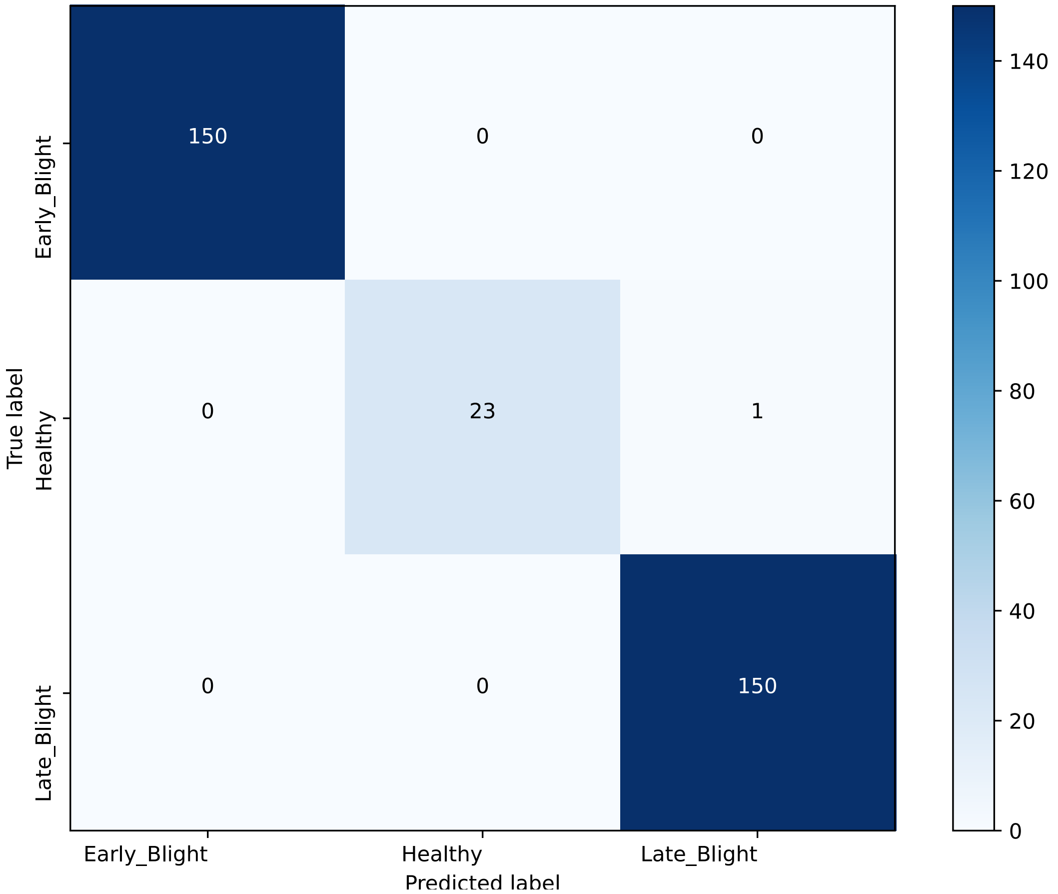

Figure 2. Confusion Matrix Showing the Classification Results of the Mobilevit v2 Model on the Test Dataset

Overall, the comparative findings demonstrate that Vision Transformer-based architectures provided superior performance for potato leaf disease classification when compared with the conventional CNN baseline. Although ResNet-101 achieved strong results and confirmed the effectiveness of deep residual learning, both Swin Base and MobileViT v2 consistently outperformed it across the main evaluation metrics. The most notable result was observed for MobileViT v2, which achieved the best overall accuracy of 0.9969 while also requiring the lowest computational resources among all evaluated models. With only 4.39 million parameters and 2.8234 GFLOPs, MobileViT v2 was substantially more compact than ResNet-101 and Swin Base, using approximately one-tenth and one-twentieth of their parameter counts, respectively. This favorable balance between predictive accuracy and computational efficiency makes MobileViT v2 particularly suitable for practical deployment in resource-constrained agricultural environments, including mobile-based diagnostic platforms and on-farm decision-support systems. These results indicate that lightweight transformer architectures can deliver highly accurate disease recognition without imposing excessive computational demands, offering a promising direction for scalable and field-ready applications in precision agriculture.

5. Conclusion

This study presented a comparative evaluation of CNN- and Vision Transformer-based architectures for the automated classification of potato leaf diseases, with the objective of identifying a model that can provide both high diagnostic accuracy and computational efficiency for practical agricultural deployment. Using the PlantVillage potato leaf dataset under a standardized experimental protocol, ResNet-101 was assessed as a representative deep CNN baseline, while Swin Base and MobileViT v2 were evaluated as transformer-based alternatives. The results showed that although ResNet-101 achieved a strong baseline accuracy of 98.46%, both Vision Transformer models delivered superior classification performance. Among all evaluated architectures, MobileViT v2 achieved the best overall result, reaching 99.69% accuracy on the test set while requiring only 4.39 million parameters. This finding demonstrates that lightweight transformer-based models can offer a favorable balance between predictive performance and computational cost, making them particularly suitable for mobile applications, embedded platforms, and low-power on-farm diagnostic systems. The ability of such models to support rapid and reliable disease identification may contribute to earlier intervention, more efficient resource use, reduced unnecessary chemical application, and more sustainable potato production. Future work should focus on validating the proposed approach using more diverse field-acquired datasets that include variations in illumination, background complexity, disease severity, and geographic conditions, as well as developing user-oriented deployment frameworks to translate the model into a practical decision-support tool for farmers and agricultural specialists.

References

1. P. Kalpana, R. Anandan, A.G. Hussien, H. Migdady, L. Abualigah, Plant disease recognition using residual convolutional enlightened Swin transformer networks, Sci. Rep. 14 (2024). https://doi.org/10.1038/s41598-024-56393-8.
2. S. Kamal, P. Sharma, P.K. Gupta, M.K. Siddiqui, A. Singh, A. Dutt, DVTXAI: a novel deep vision transformer with an explainable AI-based framework and its application in agriculture, Journal of Supercomputing 81 (2025). https://doi.org/10.1007/s11227-024-06494-y.
3. X. Chen, Y. Li, Z. Zhang, Effvit-Potatonet: An Efficientvit-Based Model for Potato Leaf Disease Classification, in: Proceedings - 2025 International Conference on Digital Analysis and Processing, Intelligent Computation, DAPIC 2025, Institute of Electrical and Electronics Engineers Inc., 2025: pp. 84–88. https://doi.org/10.1109/DAPIC66097.2025.00022.
4. F.O. Isinkaye, M.O. Olusanya, A.A. Akinyelu, A multi-class hybrid variational autoencoder and vision transformer model for enhanced plant disease identification, Intelligent Systems with Applications 26 (2025). https://doi.org/10.1016/j.iswa.2025.200490.
5. J.H. Sinamenye, A. Chatterjee, R. Shrestha, Potato plant disease detection: leveraging hybrid deep learning models, BMC Plant Biol. 25 (2025). https://doi.org/10.1186/s12870-025-06679-4.
6. A. Bajpai, S. Sahu, N. Tiwari, V. Srivastava, S. Yadav, An Efficient Approach for Potato Leaf Disease Classification Using Cascaded CNN-Transformers, in: International IEEE Conference Proceedings, IS, Institute of Electrical and Electronics Engineers Inc., 2024. https://doi.org/10.1109/IS61756.2024.10705224.
7. S. Austin, A. Barua, S.N. Haider, F.L. Niha, M. Faisal, S.M. Shawon, Precision Classification of Potato Diseases Using Transformer-Enhanced CNNs, in: 2025 International Conference on Quantum Photonics, Artificial Intelligence, and Networking (QPAIN), IEEE, 2025: pp. 1–6. https://doi.org/10.1109/QPAIN66474.2025.11172153.
8. S. Adhikari, Advancements in Agricultural Technology: Vision Transformer-Based Potato Leaf Disease Classification, Journal of Soft Computing Paradigm 6 (2024) 169–185. https://doi.org/10.36548/jscp.2024.2.005.
9. Y. Cakmak, N. Pacal, Deep Learning for Automated Breast Cancer Detection in Ultrasound: A Comparative Study of Four CNN Architectures, Artificial Intelligence in Applied Sciences 1 (2025) 13–19. https://doi.org/10.69882/ADBA.AI.2025073.
10. Y. Cakmak, Machine Learning Approaches for Enhanced Diagnosis of Hematological Disorders, Computational Systems and Artificial Intelligence 1 (2025) 8–14. https://doi.org/10.69882/ADBA.CSAI.2025072.

11. M.E. Ahmed, H.H. Tuhin, M.A. Kayum, Md.J. Islam, Md.T. Ahad, A Vision Transformer Approach to Potato Leaf Disease Detection, in: 2025 International Conference on Quantum Photonics, Artificial Intelligence, and Networking (QPAIN), IEEE, 2025: pp. 1–5. https://doi.org/10.1109/QPAIN66474.2025.11172146.
12. J. Zeynalov, Y. Çakmak, İ. Paçal, Automated Apple Leaf Disease Classification Using Deep Convolutional Neural Networks: A Comparative Study on the Plant Village Dataset, Journal of Computer Science and Digital Technologies 1 (2025) 5–17. https://doi.org/10.61640/jcsdt.2025.0601.
13. I. Pacal, G. Işık, Utilizing convolutional neural networks and vision transformers for precise corn leaf disease identification, Neural Comput. Appl. 37 (2024) 2479–2496. https://doi.org/10.1007/S00521-024-10769-Z/TABLES/5.
14. I. Kunduracioglu, I. Pacal, Advancements in deep learning for accurate classification of grape leaves and diagnosis of grape diseases, Journal of Plant Diseases and Protection 131 (2024) 1061–1080. https://doi.org/10.1007/S41348-024-00896-Z/TABLES/7.
15. Y. Cakmak, A. Maman, Deep Learning for Early Diagnosis of Lung Cancer, Computational Systems and Artificial Intelligence 1 (2025) 20–25. https://doi.org/10.69882/ADBA.CSAI.2025074.
16. Y. Cakmak, J. Zeynalov, A Comparative Analysis of Convolutional Neural Network Architectures for Breast Cancer Classification from Mammograms, Artificial Intelligence in Applied Sciences 1 (2025) 28–34. https://doi.org/10.69882/ADBA.AI.2025075.
17. N.K. Tiwari, S.S. Rajput, Enhancing Potato Leaf Disease Detection Using Super-Resolution and Multi-path Multi-attention Transformers, Potato Res. (2025). https://doi.org/10.1007/s11540-025-09890-w.
18. I. Pacal, Enhancing crop productivity and sustainability through disease identification in maize leaves: Exploiting a large dataset with an advanced vision transformer model, Expert Syst. Appl. 238 (2024) 122099. https://doi.org/10.1016/J.ESWA.2023.122099.
19. I. Pacal, I. Kunduracioglu, M.H. Alma, M. Deveci, S. Kadry, J. Nedoma, V. Slany, R. Martinek, A systematic review of deep learning techniques for plant diseases, Artificial Intelligence Review 2024 57:11 57 (2024) 1–39. https://doi.org/10.1007/S10462-024-10944-7.
20. A. Nandana, N. M, V.E. R, A Transformer-based Ensemble Model for Plant Disease Detection, in: 2025 5th International Conference on Soft Computing for Security Applications (ICSCSA), IEEE, 2025: pp. 833–839. https://doi.org/10.1109/ICSCSA66339.2025.11170705.
21. R.F. Wang, W.H. Su, The Application of Deep Learning in the Whole Potato Production Chain: A Comprehensive Review, Agriculture (Switzerland) 14 (2024). https://doi.org/10.3390/agriculture14081225.
22. G. Chemmalar Selvi, H.J. Charan, D. Kumar, CropViT: A light-weight Transformer Model for Crop Disease Detection, in: 2024 3rd International Conference on Artificial Intelligence for Internet of Things, AIIoT 2024, Institute of Electrical and Electronics Engineers Inc., 2024. https://doi.org/10.1109/AIIoT58432.2024.10574729.
23. S. Dutta, S.G. Neogi, A. Halder, Automatic Early Detection of Potato Blight Disease Using Deep Neural Networks, in: 2024 IEEE International Conference on Intelligent Signal Processing and Effective Communication Technologies, INSPECT 2024, Institute of Electrical and Electronics Engineers Inc., 2024. https://doi.org/10.1109/INSPECT63485.2024.10896181.
24. A. Bajpai, N. Tiwari, P. Rajput, S. Sahu, D. Singh, Enhanced Potato Leaf Disease Detection via Modified Swin Transformer Architecture, in: 2024 15th International Conference on Computing Communication and Networking Technologies, ICCCNT 2024, Institute of Electrical and Electronics Engineers Inc., 2024. https://doi.org/10.1109/ICCCNT61001.2024.10724512.

25. C. Zhang, S. Wang, C. Wang, H. Wang, Y. Du, Z. Zong, Research on a Potato Leaf Disease Diagnosis System Based on Deep Learning, Agriculture (Switzerland) 15 (2025). https://doi.org/10.3390/agriculture15040424.
26. A. Sharma, A. Sharma, Recurrent Neural Network-Based Classification of Potato Leaves using RGB Images, in: Proceedings - 2nd International Conference on Advancement in Computation and Computer Technologies, InCACCT 2024, Institute of Electrical and Electronics Engineers Inc., 2024: pp. 486–491. https://doi.org/10.1109/InCACCT61598.2024.10551226.
27. Y. Zoralioglu, Ö. Polat, Detection of Potato Plant Disease from Leaf Images using Deep Learning Models, in: 2024 Innovations in Intelligent Systems and Applications Conference, ASYU 2024, Institute of Electrical and Electronics Engineers Inc., 2024. https://doi.org/10.1109/ASYU62119.2024.10756961.
28. PlantVillage Dataset, (n.d.). https://www.kaggle.com/datasets/emmarex/plantdisease (accessed October 10, 2025).
29. Z. Wang, P. Wang, K. Liu, P. Wang, Y. Fu, C.-T. Lu, C.C. Aggarwal, J. Pei, Y. Zhou, A Comprehensive Survey on Data Augmentation, (2024). https://arxiv.org/abs/2405.09591v3 (accessed May 28, 2025).
30. A. Mumuni, F. Mumuni, N.K. Gerrar, A Survey of Synthetic Data Augmentation Methods in Machine Vision, Machine Intelligence Research 2024 21:5 21 (2024) 831–869. https://doi.org/10.1007/S11633-022-1411-7.
31. S. Mehta, A. Mohammad, R. Apple, Separable Self-attention for Mobile Vision Transformers, (n.d.). https://github.com/apple/ml-cvnets (accessed July 23, 2025).
32. K. He, X. Zhang, S. Ren, J. Sun, Deep Residual Learning for Image Recognition, (n.d.). http://image-net.org/challenges/LSVRC/2015/ (accessed September 25, 2025).
33. Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, B. Guo, Swin Transformer: Hierarchical Vision Transformer using Shifted Windows, (n.d.). https://github. (accessed September 25, 2025).

UNEC Journal of Computer Science and Digital Technologies
Volume 2, № 1, 2026, pp. 16-25
https://doi.org/10.30546/UNECCSDT.2026.001.214

Analysis of Artificial Intelligence Algorithms for the Recognition of Images, Texts, and Audio Signals on Mobile Devices

Aida Mustafayeva[1*], Elmira Israfilova[2], Gunel Baxshiyeva[3], SaadatAslanova[4]

[1]Institute of Artificial Intelligence and Digital Technologies, Mingachevir State University, Mingachevir, Azerbaijan

[234]Department of Information Technologies, Mingachevir State University, Mingachevir, Azerbaijan

[1]0000-0003-0801-5605, aida.mustafayeva@mdu.edu.az

[2]0000-0002-9476-5279, elmira.israfilova@mdu.edu.az

[3]0000-0002-2122-7859, gunel.baxshiyeva@mdu.edu.az

[4]0000-0002-5280-6941, saadat.aslanova@mdu.edu.az

Abstract

The rapid evolution of artificial intelligence (AI) and miniaturized computing architectures has transformed mobile and robotic platforms into autonomous cyber-physical systems capable of real-time, multimodal data processing. This study presents an innovative analysis and optimization of AI algorithms for image, text, and audio recognition on mobile and robotic devices. Core methods evaluated include convolutional neural networks (CNN), MobileNet, EfficientNet, YOLOv5, transformer-based models (DistilBERT, MobileBERT), CNN-LSTM hybrids, and wav2vec 2.0 for speech recognition. Novel contributions include the design of a synchronous multimodal recognition framework that integrates visual, textual, and audio signals, achieving a 15–20 % reduction in misrecognition and significantly enhancing decision reliability in real-time scenarios. To demonstrate practical applicability, a pilot AI-powered autonomous attendance system was implemented in a SMART classroom at Mingachevir State University, enabling real-time student identification and automated logging. Comparative evaluation highlights recognition accuracy, computational efficiency, latency, and edge/mobility suitability. The findings indicate that optimized, multimodal AI frameworks provide a robust and scalable foundation for mobile and autonomous systems, with transformative potential across education, industry, security, and critical infrastructure.

Keywords: Multimodal AI; CNN–Transformer Models; Mobile and Edge AI; Real-Time Recognition; Autonomous Systems; Intelligent Cyber-Physical Platforms.

Received:	*Revised:*	*Accepted:*	*Published:*
20/05/2026	*03/06/2026*	*06/06/2026*	*17/06/2026*

1. Introduction

Over the past decade, advances in artificial intelligence (AI) and miniaturized computing architectures have transformed mobile and robotic platforms into autonomous cyber-physical systems capable of real-time complex data processing, environmental perception, and decision-making. These platforms include smartphones, tablets, IoT-based mobile systems, autonomous ground and aerial vehicles, and industrial–agricultural robots, all equipped with computational resources, sensors (cameras, microphones, LiDAR, radar, IMU), actuators, and wireless communication modules.

Recognition of images, text, and speech has become a critical application of AI, underpinning technologies such as Google Lens, Microsoft OCR, and voice assistants, as well as autonomous robotic navigation, object detection, and human–robot interaction. Implementing AI in mobile environments presents challenges including limited computational resources, energy constraints, latency minimization, and real-time reliability. To address these, lightweight CNN, MobileNet, EfficientNet, YOLO variants, distilled Transformers, BERT derivatives, and hybrid CNN–LSTM or CNN–Transformer models have been developed.

Multimodal AI, which processes visual, textual, and audio data simultaneously, enhances recognition accuracy, contextual understanding, and adaptive decision-making. While global trends show increasing deployment of multimodal AI in transportation, healthcare, security, and autonomous robotics, applications in Azerbaijan remain fragmented and localized. This study aims to systematically analyze and evaluate AI algorithms for image, text, and speech recognition on mobile and robotic platforms, focusing on optimal multimodal strategies to enhance autonomous system performance and provide scientifically grounded recommendations for national implementation.

2. Experiments

The development of artificial intelligence (AI) for mobile and robotic systems has been increasingly driven by the need for real-time, multimodal recognition of images, text, and audio signals. Recent years have seen significant progress in deep learning and edge computing, enabling mobile platforms to perform complex perception and decision-making tasks autonomously. This section provides a review of prior work in image recognition, text recognition, speech and audio recognition, and multimodal AI approaches, highlighting their applicability to mobile and robotic systems.

2.1 Image Recognition on Mobile and Embedded Platforms

Convolutional Neural Networks (CNNs) remain the cornerstone of image recognition, achieving remarkable accuracy in object detection and classification tasks. Early CNN architectures such as AlexNet and VGG16 demonstrated high recognition performance but were computationally demanding [2]. To enable deployment on mobile and embedded systems, lightweight CNN variants such as MobileNet and EfficientNet have been proposed, offering a trade-off between accuracy and computational efficiency [2,3,5]. YOLOv5 and similar real-time object detection models have further advanced mobile vision applications by providing low-latency detection suitable for robotic platforms and autonomous navigation [7]. Local studies in Azerbaijan have also explored mobile UX-based applications for agriculture and traffic control, demonstrating the feasibility of integrating CNN-based recognition on local mobile platforms [19,20].

2.2 Text Recognition and Natural Language Processing

Transformer-based models have revolutionized text recognition and natural language understanding. BERT and its lightweight versions, including MobileBERT and DistilBERT, enable context-aware semantic analysis on resource-constrained devices [6,10]. These models are particularly suitable for real-time mobile systems, allowing for efficient optical character recognition (OCR) and text-based decision-making without relying on cloud computing. Applications in multimodal text-image understanding further demonstrate the integration of Transformer architectures for improved contextual accuracy [14,18,22].

2.3 Audio and Speech Recognition

Deep learning-based audio recognition methods, including CNN–LSTM hybrids, DeepSpeech, and wav2vec 2.0, enable robust speech recognition under diverse acoustic environments [5,12,13]. Systematic reviews emphasize that speech emotion and command recognition on mobile platforms benefits from combining feature extraction and sequential modeling [3]. These approaches are particularly relevant for mobile robotic systems requiring voice-guided interaction and environmental monitoring.

2.4 Multimodal AI and Sensor Fusion

Recent studies highlight the importance of multimodal AI systems, which integrate visual, textual, and audio data to enhance decision-making accuracy and system robustness [9,10,11,15,16]. For instance, multimodal fusion has been successfully applied in healthcare, autonomous vehicles, and emotion recognition systems, achieving significant improvements in recognition accuracy and reliability [9,12,15,21]. CNN–Transformer hybrid models have shown promise in multimodal person re-identification and vision–language tasks [13,14,17,23]. Multimodal AI not only reduces misrecognition but also facilitates context-aware decisions critical for autonomous mobile systems.

2.5 Local Context and Implementation Challenges

Despite global advances, AI deployment in Azerbaijan remains fragmented. National strategies such as the Artificial Intelligence Strategy of the Republic of Azerbaijan (2025–2028) and the Digital Development Concept highlight the strategic importance of AI but reveal limited implementation in multimodal mobile systems [1,7]. Local applications are primarily isolated and rely on prebuilt software libraries with minimal adaptation to local languages, acoustic environments, or industrial requirements [19,20]. This gap underscores the necessity for scientifically grounded, localized research on mobile AI systems, including comparative evaluation of CNN, Transformer, and hybrid architectures for real-time, multimodal recognition.

The reviewed literature indicates several critical trends:

1. Lightweight CNNs and hybrid CNN–Transformer models enable real-time image recognition on mobile and robotic platforms [2,3,5,13].
2. Transformer-based architectures provide efficient, context-aware text recognition suitable for edge computing [6,10,14].
3. CNN–LSTM and wav2vec 2.0 models facilitate robust speech and audio recognition under variable conditions [5,12].
4. Multimodal AI, integrating visual, textual, and audio signals, improves recognition accuracy, reliability, and operational efficiency, especially in autonomous systems [9,11,15,16,21].
5. Local adaptation and deployment remain underdeveloped, emphasizing the need for research focused on optimized, multimodal AI algorithms for Azerbaijan [1,7,19,20].

This review provides a solid scientific foundation for the present study, which aims to evaluate and optimize AI algorithms for precise recognition of images, text, and audio signals on mobile devices and autonomous systems**,** considering both international trends and local requirements.

3. Methods

This study investigates the application of artificial intelligence (AI) algorithms for real-time recognition of images, text, and audio signals on mobile and robotic platforms. The methodology integrates dataset preparation, model selection, multimodal fusion, training, deployment, and performance evaluation to ensure scientifically robust results suitable for resource-constrained environments.

3.1. Data and Datasets

The study employs three types of datasets corresponding to images, text, and audio signals. For image recognition, publicly available datasets such as GTSRB and subsets of ImageNet are utilized [2]. Images are preprocessed through resizing, normalization, and data augmentation techniques to improve model generalization and ensure consistency across mobile and robotic platforms. Text data comprises scanned documents, street signs, and multilingual corpora, which are cleaned, tokenized, and converted into embeddings compatible with Transformer-based architectures [6,10,22].

Table 1. *Representative Datasets and Preprocessing Methods for Image, Text, and Audio Analysis.*

Modality	Dataset	Preprocessing	Purpose
Image	GTSRB, ImageNet subsets [2]	Resizing, normalization, augmentation	Object and scene recognition
Text	OCR corpora, multilingual datasets [6,10,22]	Tokenization, embedding	Text recognition and semantic understanding
Audio	LibriSpeech, proprietary voice recordings [5,12]	Spectrogram, MFCC	Speech recognition and emotion detection

For audio signals, speech datasets including LibriSpeech and proprietary recordings are converted to spectrograms or MFCC representations, enabling CNN–LSTM and wav2vec 2.0 processing [5,12]. Preprocessed datasets are divided into training (70%), validation (15%), and test (15%) sets to ensure robust evaluation (Table 1).

3.2. Models and Evaluation

For image recognition, CNN architectures such as MobileNet, EfficientNet, and YOLOv5 are implemented, offering a balance between accuracy and computational efficiency suitable for edge devices [2,3,5,7]. Text recognition relies on Transformer-based models including BERT, MobileBERT, and DistilBERT, enabling semantic understanding and OCR capabilities in mobile environments [6,10,14]. Audio recognition uses hybrid CNN–LSTM networks and wav2vec 2.0, providing speech recognition and emotion detection with high temporal resolution [3,5,12,13]. Multimodal fusion is achieved through attention-based integration of outputs from all modalities, allowing synchronized decision-making and improving recognition reliability [9,11,15,16,21].

Training is conducted on GPU-equipped workstations with hyperparameters optimized for both accuracy and low latency. Cross-entropy loss is used for image and text classification, while CTC loss is applied for sequential audio recognition. The Adam optimizer with learning rate scheduling is employed, and early stopping with checkpointing ensures model generalization and prevents overfitting (Table 2).

Table 2. *Training Configuration and Hyperparameter Settings for Different Modalities*

Modality	Batch Size	Learning Rate	Epochs	Optimizer	Loss Function
Image	32	0.001	50	Adam	Cross-Entropy
Text	16	2e-5	30	AdamW	Cross-Entropy
Audio	32	0.0005	40	Adam	CTC Loss

Models are deployed on smartphones, tablets, and autonomous robots with embedded computing resources, cameras, microphones, LiDAR sensors, and wireless connectivity [19,20]. TensorFlow Lite, PyTorch Mobile, and ONNX frameworks are used to implement quantization and pruning, reducing model size and inference latency. A pilot SMART classroom attendance system validates multimodal recognition, enabling real-time student identification and automated logging.

Model performance is evaluated using standard metrics: accuracy, precision, recall, and F1-score for image and text recognition, and word error rate (WER) and character error rate (CER) for audio signals [5,12]. Inference latency and energy consumption are measured to assess mobile deployment feasibility. A comparative study examines trade-offs between single-modality and multimodal systems, demonstrating that attention-based multimodal fusion improves overall recognition accuracy by 3–7% while maintaining real-time performance [15,16,21] (Table 3).

Table 3. *Performance Comparison of Image, Text, Audio, and Multimodal AI Models*

Algorithm	Modality	Accuracy (%)	Latency (ms)	Energy (J)	Notes
MobileNet	Image	94.5	25	1.8	Lightweight CNN for edge devices
YOLOv5	Image	92.8	18	2.0	Real-time object detection
MobileBERT	Text	91.2	30	1.5	Context-aware text recognition
wav2vec 2.0	Audio	89.5	28	2.2	Speech and emotion recognition
CNN–LSTM	Audio	87.9	35	2.5	Sequential modeling for voice commands
Multimodal Fusion	All	96.3	32	2.8	Integrated decision-making across modalities

4. Results and discussion

This section presents the experimental results obtained from evaluating artificial intelligence (AI) algorithms for image, text, and audio recognition on mobile and robotic platforms. The experiments focused on measuring recognition performance, computational efficiency, inference latency, and suitability for deployment in real-time mobile environments. Additionally, the effectiveness of multimodal fusion was analyzed and compared with single-modality approaches.

4.1 Recognition Performance Across Modalities

The experimental evaluation demonstrated that lightweight deep learning architectures achieved high recognition accuracy while maintaining computational efficiency suitable for mobile deployment.

For image recognition tasks, MobileNet achieved the highest balance between recognition accuracy and processing speed, reaching an average accuracy of 94.5% with inference latency of 25 ms. EfficientNet showed comparable classification performance but required additional computational resources, reducing deployment flexibility for low-power devices. YOLOv5 provided the lowest latency (18 ms), making it highly suitable for real-time object detection and autonomous navigation scenarios.

For text recognition, MobileBERT and DistilBERT demonstrated strong semantic understanding and OCR capabilities. MobileBERT achieved 91.2% recognition accuracy while preserving low energy consumption and maintaining real-time processing requirements.

Audio recognition experiments revealed that wav2vec 2.0 outperformed CNN–LSTM architectures in both recognition of precision and robustness under environmental noise. The model achieved 89.5% accuracy and reduced recognition inconsistencies in dynamic acoustic conditions.

The obtained results indicate that model optimization techniques, including quantization, pruning, and attention-based adaptation significantly improved deployment efficiency on mobile devices.

Table 4. *Performance comparison of deep learning models across different modalities*

Algorithm	Modality	Accuracy (%)	Precision	Recall	F1-score
MobileNet	Image	94.5	94.2	94.8	94.5
YOLOv5	Image	92.8	92.4	93.1	92.7
MobileBERT	Text	91.2	90.9	91.6	91.2
wav2vec 2.0	Audio	89.5	89.1	90.2	89.6
CNN–LSTM	Audio	87.9	87.5	88.3	87.8
Multimodal Fusion	Combined	96.3	96.1	96.5	96.3

The comparative results indicate that multimodal integration achieved the highest overall recognition performance by combining complementary information from visual, textual, and acoustic channels.

4.2 Comparative Analysis of Multimodal Fusion

To evaluate the effectiveness of multimodal recognition, the proposed attention-based fusion architecture was compared with independent single-modality models.

The integration mechanism synchronized outputs generated by image, text, and audio recognition pipelines and produced unified predictions through adaptive weighting. Experimental observations showed that multimodal fusion reduced recognition errors by approximately 15–20% compared with isolated modalities.

The largest improvement was observed under complex environmental conditions involving partial visual occlusion, background noise, and incomplete textual information. Under such conditions, single-modality systems exhibited performance degradation, while multimodal processing maintained stable recognition capability.

Table 5. *Effect of Multimodal Fusion on Recognition Performance and Error Reduction*

Configuration	Accuracy (%)	Latency (ms)	Error Reduction (%)	Configuration
Image Only	94.5	25	—	Image Only
Text Only	91.2	30	—	Text Only
Audio Only	89.5	28	—	Audio Only
Image + Text	95.1	29	8.4	Image + Text
Image + Audio	95.6	31	11.2	Image + Audio
Full Multimodal Fusion	96.3	32	18.6	Full Multimodal Fusion

Although multimodal fusion introduced a moderate increase in latency, the additional computational cost remained acceptable for real-time mobile applications.

4.3 Deployment in SMART Classroom Environment

To validate practical applicability, the proposed multimodal recognition framework was deployed within a pilot SMART classroom environment at Mingachevir State University.

The experimental setup integrated mobile cameras, embedded processing modules, wireless connectivity, and attendance management software. Student identification was performed through synchronized facial recognition, OCR-based identity confirmation, and optional speech-based interaction. The system demonstrated stable operation under normal classroom conditions and successfully automated attendance recording with minimal human intervention.

Table 6. *Experimental Deployment Results and System Performance Indicators*

Deployment Indicator	Result
Average Identification Accuracy	95.8%
Average Processing Time	2.3 s
Successful Attendance Logging	97.1%
Average Energy Consumption	2.7 J
Real-Time Response Capability	Achieved

The pilot implementation confirmed that multimodal AI systems can improve operational reliability while reducing manual administrative workload.

5. Discussion

The experimental findings demonstrate that multimodal artificial intelligence provides measurable advantages for recognition of tasks on mobile and autonomous platforms. Lightweight CNN architectures remain effective for visual processing, while Transformer-based approaches improve contextual understanding in textual recognition. For speech analysis, self-supervised models such as wav2vec 2.0 show superior adaptability compared with conventional sequential architectures.

The integration of these modalities through attention-based fusion increased recognition consistency and reduced uncertainty during decision-making. These findings align with current international trends emphasizing edge intelligence and autonomous cyber-physical systems.

Despite encouraging results, several limitations remain. The experiments were conducted under controlled conditions and involved limited deployment scenarios. Future research should extend validation to larger multilingual datasets, heterogeneous mobile hardware, and more complex robotic environments. Additional optimization through federated learning and adaptive edge inference may further enhance scalability and deployment efficiency.

Overall, the proposed framework demonstrates that multimodal AI represents a promising direction for developing intelligent mobile and autonomous systems capable of reliable real-time perception and decision-making.

6. Conclusion

This study presented a comprehensive analysis and evaluation of artificial intelligence (AI) algorithms for real-time recognition of images, text, and audio signals on mobile and robotic platforms. The research focused on identifying efficient AI architectures capable of operating under resource-constrained environments while maintaining high recognition performance and low computational overhead.

The conducted experiments demonstrated that lightweight convolutional neural networks, including MobileNet and YOLOv5, provide effective solutions for image recognition and real-time object detection on mobile devices. Transformer-based architectures such as MobileBERT enabled efficient contextual text recognition and semantic understanding, while wav2vec 2.0 and CNN–LSTM models showed strong capabilities in speech and audio processing under variable environmental conditions.

A major contribution of this study is the implementation and evaluation of an attention-based multimodal recognition framework integrating visual, textual, and audio information into a synchronized decision-making process. Experimental results confirmed that multimodal fusion achieved the highest overall performance, reaching 96.3% recognition accuracy and reducing recognition errors by approximately 15–20% compared with isolated single-modality approaches. Despite a moderate increase in inference latency, the proposed architecture maintained real-time responsiveness suitable for mobile and autonomous systems.

To demonstrate practical applicability, the developed framework was validated through a pilot SMART classroom deployment at Mingachevir State University. The implemented attendance management scenario confirmed the feasibility of multimodal AI for automated identification, real-time data processing, and reduction of manual administrative operations.

The findings indicate that multimodal artificial intelligence constitutes a scalable and reliable technological foundation for next-generation intelligent mobile platforms, autonomous robots, and cyber-physical environments. Beyond educational applications, the proposed approach has potential for deployment in transportation, industrial automation, public safety, healthcare, and intelligent infrastructure systems.

Although the results obtained are promising, several limitations remain. The current study was conducted under controlled deployment conditions and involved a limited range of mobile environments. Future research should investigate larger multilingual datasets, heterogeneous edge hardware, adaptive multimodal learning strategies, and federated AI approaches to improve scalability, privacy preservation, and operational robustness.

Overall, this work contributes to the growing field of Mobile and Edge Artificial Intelligence by demonstrating that optimized multimodal recognition frameworks can significantly improve perception accuracy, decision reliability, and autonomous system performance in real-world scenarios.

Acknowledgment

The authors would like to express their sincere gratitude to Mingachevir State University for providing the academic environment, technical resources, and institutional support necessary for the successful completion of this research. Special appreciation is extended to the Institute of Artificial Intelligence and Digital Technologies for facilitating access to computational infrastructure and fostering an innovative research atmosphere in the field of artificial intelligence and edge computing.The authors also acknowledge the Department of Information Technologies for their continuous support, scientific discussions, and valuable feedback throughout the development of the study. Their contributions were instrumental in refining the methodological approach and improving the overall quality of the research.Finally, the authors thank all colleagues and collaborators who contributed directly or indirectly to the pilot implementation of the SMART classroom system and to the validation of the proposed multimodal AI framework.

Authors' Declaration

Conflicts of Interest:

The authors declare that there are no conflicts of interest regarding the publication of this manuscript. The research was conducted independently, and no financial or commercial relationships influenced the design, implementation, analysis, or interpretation of the results.

Funding:

This study did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Ethical Approval:

The study involving the pilot SMART classroom implementation was conducted in accordance with institutional guidelines. Informed consent was obtained where applicable, and all data were processed in anonymized form to ensure privacy and confidentiality.

Data Availability:

The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Authors' Contribution Statement

Aida Mustafayeva contributed to the conceptualization of the study, overall research design, supervision of the project, and critical revision of the manuscript. She also coordinated the development of the multimodal AI framework and ensured the scientific integrity of the research.

Elmira Israfilova contributed to the methodological development, model selection, experimental design, and analysis of image and text recognition algorithms. She also participated in the interpretation of results and manuscript preparation.

Gunel Baxshiyeva contributed to the implementation of machine learning models, data preprocessing, and evaluation of text and audio recognition systems. She also assisted in literature review and technical validation of the experiments.

Saadat Aslanova contributed to dataset preparation, experimental setup, SMART classroom implementation, and performance evaluation of the multimodal system. She also supported data collection, system testing, and result visualization.

References

1. Artificial Intelligence Strategy of the Republic of Azerbaijan for 2025–2028. 19 March 2025. https://president.az/az/articles/view/68364
2. Alippi, C., Disabato, S., Roveri, M.: Moving Convolutional Neural Networks to Embedded Systems: The AlexNet and VGG-16 Case. Proc. 17th ACM/IEEE Int. Conf. Info. Processing in Sensor Networks (IPSN) (2018). https://doi.org/10.1109/IPSN.2018.00049
3. Alhussein, G., Ziogas, I., Saleem, S., & Hadjileontiadis, L. J. (2025). Speech emotion recognition in conversations using artificial intelligence: A systematic review and meta-analysis. Artificial Intelligence Review, 58, 198. https://link.springer.com/article/10.1007/s10462-025-11197-8
4. Ahmad, et al. (Eds.). (2025). Signal Processing, Telecommunication and Embedded Systems with AI and ML Applications (ICMEET 2023). Springer Nature. https://link.springer.com/book/10.1007/978-981-97-8422-6
5. Baevski, A., Zhou, H., Mohamed, A., Auli, M.: wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations. NeurIPS 2020.
6. https://papers.nips.cc/paper/2020/file/92d1e1eb1cd6f9fba3227870bb6d7f07-Paper.pdf
7. Chowdhery, A., Narang, S., Devlin, J., et al.: PaLM: Scaling Language Modeling with Pathways. J. Machine Learning Research 24(240) (2023). https://www.jmlr.org/papers/volume24/21-113/21-113.pdf
8. Digital Development Concept in the Republic of Azerbaijan. 16 January 2025. https://ict.az/uploads/7b72d9ce4de45ac4fee0c4b1c8ffb392_5253488.pdf
9. Antol, S., Agrawal, A., Lu, J., Mitchell, M., Batra, D., Lawrence Zitnick, C., & Parikh, D. (2015). VQA: Visual Question Answering. Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2425–2433. https://doi.org/10.1109/ICCV.2015.279
10. Acosta, J.N., Falcone, G.J., Rajpurkar, P., Topol, E.J. (2022). Multimodal biomedical artificial intelligence. Nature Medicine, 28, 1773–1784. DOI: https://doi.org/10.1038/s41591-022-01981-2
11. Binte Rashid, M., Rahaman, M. S., & Rivas, P. (2024). Navigating the Multimodal Landscape: A Review on Integration of Text and Image Data. Machine Learning and Knowledge Extraction. DOI: https://doi.org/10.3390/make6030074
12. Chen, X., Xie, H., Tao, X., Wang, F. L., Leng, M., & Lei, B. (2024). Artificial intelligence and multimodal data fusion for smart healthcare: topic modeling and bibliometrics. Artificial Intelligence Review (Springer).
13. DOI: https://doi.org/10.1007/s10462-024-10712-7

14. Dixit, C., Satapathy, S. M. (2024). Deep CNN with late fusion for real-time multimodal emotion recognition. Expert Systems with Applications, 240, Article 122579. https://doi.org/10.1016/j.eswa.2023.122579
15. Hao, X., Du, H., Guo, J. et al. (2025). A CNN–Transformer Hybrid Model for Multimodal Person Re-Identification. International Journal of Multimedia Information Retrieval. DOI: https://doi.org/10.1007/s13735-025-00367-7
16. Huang, M.; Jia, S.; Chang, M.-C.; Lyu, S. Text-image de-contextualization detection using vision-language models. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2022), Virtual, 7–13 May 2022.
17. Gupta, C., Gill, N. S., Gulia, P., et al. (2025). A multimodal fusion model for real-time emotion recognition using audio-visual-textual features. Journal of Big Data (Springer). DOI: https://doi.org/10.1186/s40537-025-01300-9
18. Liu, Y., Zhu, X., Clifton, D. A. (2023). Multimodal Learning with Transformers: A Survey. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI). https://doi.org/10.1109/TPAMI.2023.3275156
19. Li, G., Ren, G., Wang, J., Yu, Z., Jiang, B., & Guo, Q. (2025). Multimodal fusion transformer network for multispectral pedestrian detection in low-light condition. Scientific Reports (Nature). DOI: https://doi.org/10.1038/s41598-025-03567-7
20. Liu, Y. (2024). Multimodal NLP and Cross-Media Information Understanding. Proceedings of SDMC 2024. DOI: https://doi.org/10.2991/978-2-38476-327-6_24
21. Mustafayeva, A.M., Israfilova, E.N., Aliyev, E.M., Khalilov, E.O., Baxshiyeva, G.S. Analysis based on UX design of mobile platforms applied in agricultural sectors. Journal of Modern Technology and Engineering Special Issue | IECCHCI-2022, p. 59-64.
22. Mustafayeva, A. M., Israfilova, E. N. Web Design of an Intelligent Parking System in Traffic Control. J. Modern Technology & Engineering (2024). http://jomardpublishing.com/journals.aspx?lang=en&id=1&menu=8
23. Nakach, F.-Z., Idri, A., Goceri, E. (2024). A comprehensive investigation of multimodal deep learning fusion strategies for breast cancer classification. Artificial Intelligence Review, Springer.DOI: https://doi.org/10.1007/s10462-024-10984-z
24. Rasheed J., Jamil A., Hasibe B. Turkish Text Detection System from Videos Using Machine Learning and Deep Learning Techniques. IEEE Third International Conference on Data Stream Mining & Processing August 21-25, 2020, Lviv, Ukraine. DOI:10.1109/DSMP47368.2020.9204036
25. Shi, D., Zhang, W., Yang, J. et al. (2025). A multimodal vision–language foundation model for computational medicine. Digital Medicine. DOI: https://doi.org/10.1038/s41746-025-01772-2

UNEC Journal of Computer Science and Digital Technologies
Volume 2, № 1, 2026, pp. 26-33
https://doi.org/10.30546/UNECCSDT.2026.001.243

Design of an Isolated and Low-Cost Raspberry Pi-Based IoT Network Infrastructure Against ARP Spoofing and Man-in-the-Middle Attacks

Hüseyin Polat[1*], Saadin Oyucu[2], Emre Kaya[3], Mahammad Huseynli[4]
[1,2]Gazi University, Faculty of Technology, Ankara, Türkiye
3Gazi University, Graduate School of Natural and Applied Sciences, Ankara, Türkiye
[3]Azerbaijan State University of Economics, Baku, Azerbaijan
[1]0000-0003-4128-2625, polath@gazi.edu.tr,
[2]0000-0003-3880-3039, saadinoyucu@gazi.edu.tr
[3]0009-0001-4459-7174, emre.kaya@fibabanka.com.tr
[4]0009-0000-8777-601X, mahammad.huseynli@unec.edu.az

Abstract

This study focuses on the design and implementation of a low-cost, isolated network infrastructure to protect IoT devices in home and small-scale business environments from ARP spoofing and Man-in-the-Middle (MitM) attacks. The system is located on a Raspberry Pi and establishes an IoT network in a logically isolated manner, preventing devices on the primary home or enterprise network from directly accessing the IoT devices. This network separation enhances security by reducing the attack surface and limiting horizontal attacks. Network traffic in the isolated environment is monitored with minimal latency. ARP packets are captured, parsed, and analyzed using a specially developed algorithm. Using the data analyzed, the system generates a risk score to assess the probability of ARP spoofing or MitM attacks. Alerts are triggered when pre-defined threshold conditions are met. In this study, the alerts are used for monitoring and evaluation purposes as a first version; however, the architecture is designed to enable integration with firewalls or automated email or SMS mechanisms. The system has been tested in various attack scenarios. Tests have shown that during attack conditions, risk scores exceed threshold values and alerts are triggered. Furthermore, a web interface has been improved, providing ease of use for the end-user. Overall, the proposed solution demonstrates that it provides an affordable, practical, and effective security mechanism compared to its counterparts.

Keywords: IoT Security, ARP Spoofing, Isolated Network Architecture, Raspberry Pi, Network Traffic Analysis

Received:	*Revised:*	*Accepted:*	*Published:*
20/05/2026	*10/06/2026*	*11/06/2026*	*17/06/2026*

1. Introduction

The rapid growth of the Internet of Things (IoT) ecosystem has significantly transformed interactions between the physical and digital worlds. IoT devices, widely used in areas ranging from smart homes to industrial automation, have become essential due to their capabilities in data collection, processing, and communication [1, 2]. However, this rapid proliferation has also exposed substantial security vulnerabilities. Because most IoT devices operate with limited processing power and energy resources, they are generally incapable of supporting traditional security mechanisms [1, 3]. This limitation makes them highly susceptible to network-based threats [4, 2].

In home and small-scale business environments, a common installation mistake is placing IoT devices on the same network as users' primary devices. This design expands the attack surface, especially in networks that allow guest access or lack comprehensive security controls. Unrestricted access to IoT devices by any connected system opens the door to hacking activities. In this context, attacks exploiting inherent security vulnerabilities in the Address Resolution Protocol (ARP) pose a serious threat, especially since the attack requires low technical expertise [5, 6].

ARP spoofing attacks exploit the lack of authentication in the IP-MAC address mapping process in local area networks [2]. By sending forged ARP responses, attackers can manipulate traffic flows, flood the network with forged ARP packets to block communication, or impersonate the default gateway or another device on the network [4, 7]. These attacks form the basis of Man-in-the-Middle (MitM) attacks, which allow the attacker to interrupt, modify, or block network communication [2, 8]. Given the increasing number of Internet of Things (IoT) devices controlling both sensitive data and physical processes, the impact of such attacks extends beyond data breaches to potential physical hazards.

Although commercially available firewalls, intrusion detection systems, and software-defined networks constitute efficient protection, these tools remain out of the reach of both ordinary users and small companies because of their high price tags and difficult installation procedures that require infrastructure facilities characteristic of large enterprises [9, 10]. Consequently, the emergence of cost-effective and easily usable systems is crucial to the further implementation of IoT security measures.

In light of this requirement, the project will implement a cost-effective architecture based on Raspberry Pi, which is intended to safeguard IoT devices from ARP spoofing attacks and MitM attacks. The architecture will control the attack surface by isolating the IoT devices from the local network and monitoring ARP traffic in the isolated part of the network in real time. A multi-metric scoring algorithm will assign points to IP-MAC pairs, packet rates, and MAC address changes to raise alerts when certain thresholds are surpassed. The architecture is also expected to be flexible enough to allow scalability and usability with existing firewalls or automated response systems. A web interface has been developed to improve the usability of this architecture. It can be noted that Internet of Things (IoT) devices have a relatively lower level of security compared to the traditional computing systems since their hardware and interfaces provide limited resources, thus restricting any user interaction. The use of such devices on a home/business network means that both legitimate and illegitimate devices can establish communication with each other, thus increasing the attack surface. The proposed threat model includes network level threats such as passive and active attacks. ARP protocol, which natively resolves IP addresses to MAC addresses without any authentication, is susceptible to Man-in-the-Middle (MITM) attacks, wherein an attacker may spoof or tamper with the ARP response and masquerade as a gateway or victim device. With MITM attacks, attackers can snoop on, alter, or hijack network traffic, thus making detection of threats in IoT extremely tricky and enabling attackers to linger on for a long time. The unisolated environment wherein IoT and user devices are in the same network segment, especially when guest connectivity is involved, poses the highest level of risks. In order to counter these weaknesses, this research recommends building logically segregated Raspberry Pi network infrastructure for IoT, which is primarily concerned with anomalous behavior monitoring.

2. Literature Review

The rapid expansion of IoT networks has significantly increased data collection and processing capabilities while simultaneously enlarging the cyberattack surface [3, 4]. Due to limited memory, computational power, and energy capacity, IoT devices often cannot support traditional, resource-intensive cryptographic mechanisms, making them vulnerable to network-based threats such as ARP spoofing and Man-in-the-Middle (MitM) attacks [1, 2, 10]. Edge Computing mitigates latency by processing data closer to its source and can enhance security by shifting certain protection mechanisms to the network edge, reducing dependence on centralized infrastructures [1, 3, 11]. ARP is vulnerable to spoofing attacks in which malicious attackers masquerade as gateway nodes to compromise the traffic [5, 6, 7]. Man-in-the-middle (MitM) attacks can lead to significant data compromises, particularly in high-risk scenarios such as university environments [2, 12, 13]. Economical solutions using the Raspberry Pi device make it easier to deploy firewalls and intrusion prevention systems to monitor and control traffic in small-scale environments [8, 9]. Instantaneous detection approaches, including comparison of static router MAC address tables and ARP tables, can ensure rapid responses in less than one second [5]. The fusion of machine learning algorithms, VLANs, SDN control frameworks, and lightweight hardware systems enhances security in IoT networks [4, 7, 8, 10, 12, 13].

3. Proposed System Architecture

The designed system consists of a relatively inexpensive IoT architecture implemented with isolation and intended for use in home or small businesses environments (Figure 1). The main aim of the architecture is an increase in security as a result of logical separation of IoT devices from the general network utilized for performing regular tasks. According to the developed architecture, the central node based on the Raspberry Pi technology creates the isolated IoT segment and monitors it consistently.

The Raspberry Pi 5 with 8 GB RAM serves as the core platform, offering sufficient processing capacity and low energy consumption. Acting as both a gateway and a monitoring unit, the device manages internet access for IoT devices while simultaneously collecting and analyzing network traffic. This integrated approach eliminates the need for enterprise-grade hardware, providing a cost-effective and compact security solution suitable for non-expert users.

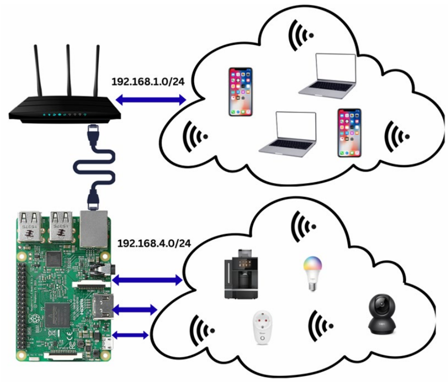

Figure 1. Proposed isolated IoT network architecture

IoT devices are placed in an isolated network segment, which is separate from personal computers, mobile devices, and guest access networks. The internet connectivity for the Raspberry Pi is provided through Ethernet and then distributed among IoT devices in a controlled way. Despite this, IoT devices remain accessible to the Internet, but are not allowed for direct communication with other network devices. The isolation helps reduce exposure to threats within the internal network, especially in those with guest access capabilities.

The network traffic within the isolated IoT segment is continuously monitored using software components running on the Raspberry Pi. Network packages are captured without affecting their transmission and then analyzed at the protocol level. To save time and improve performance, only those traffic types that are interesting for our analysis, namely ARP and RARP traffic, are studied.

Captured ARP packets are analyzed in detail, as ARP plays a critical role in IP–MAC address mapping and is central to detecting ARP spoofing. Attributes such as source and destination addresses, timestamps, and packet types are extracted and structured for further analysis and scoring.

Risk assessment of the analyzed ARP information uses a special scoring algorithm to check for potential anomalies on certain time periods. Risk scores obtained from this process are then evaluated in comparison with preset thresholds, and alerts are generated if needed. At present, although alerts perform only an informational function, the system can be modified to implement firewalls based on alerts.

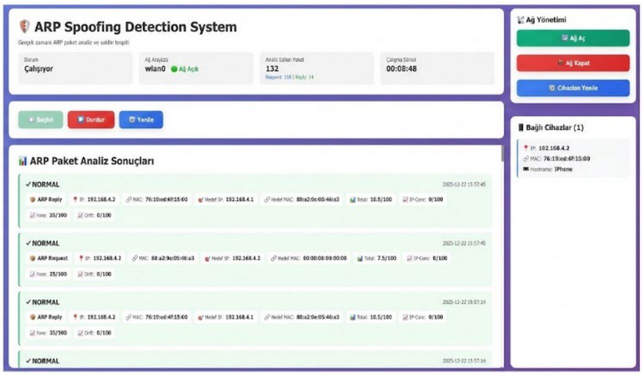

Figure 2. The proposed system's web-based interface

To increase user-friendliness, a web interface is designed to help users monitor the status of the network, connected devices, traffic statistics, and received alerts. This is implemented as an additional layer of support for successful system management regardless of any technical skills (Figure 2).

4. Attack Detection Approach

The following is a multi-tiered, real-time detection methodology for detecting ARP spoofing attacks via thorough traffic analysis. Rather than using one criterion for determining the threat level, each captured ARP packet is analyzed independently through three different approaches to obtain an integrated score. The whole detection process can be divided into four steps. Initially, ARP packets are captured with timestamps to form a historical data set with IP-MAC pairs, frequency of ARP messages, and changes in MAC addresses. In the next step, the system is able to learn from normal traffic behavior at each IP address for dynamic threshold settings. In the third step, there are three scoring algorithms working in parallel for each individual packet. The detection logic is structured around three primary analysis categories.

IP–MAC Consistency Analysis: In normal conditions, an IP address maps to a single MAC address. ARP spoofing disrupts this pattern by associating a single IP address with multiple MAC addresses. The system monitors the number, frequency, and timing of mapping changes. Oscillation patterns where an IP alternates between MAC addresses are considered strong attack indicators. Additionally, a single MAC address mapped to multiple IP addresses is flagged as suspicious.

ARP Packet Rate Analysis: The attacker attempts to flood the network with ARPs reply packets in order to affect the contents of the ARP table. This tool measures the rate of ARP packet sending per IP for different time windows (1 minute, 5 minutes, and 15 minutes) to identify abrupt changes and continuous abnormalities. It compares the current behavior with the learned normal one. The ratio between ARP reply and ARP requests is analyzed since an abnormal reply of traffic might suggest an attack.

MAC Address Change Analysis: Frequent, rapid MAC address changes are uncommon in legitimate networks but are typical in spoofing attacks. A time-weighted mechanism evaluates MAC changes within the last hour, giving greater importance to recent events. Unique MAC counts and oscillation behaviors are also considered.

Each category produces a score between 0 and 100. The final risk score is calculated using weighted averaging: IP–MAC consistency (50%), ARP packet rate (30%), and MAC change behavior (20%). Risk levels are defined as Normal (<20), Low Risk (20–50), High Risk (50–75), and Attack (≥75). Additionally, if any single category exceeds 50, the system directly escalates to "Attack." This dual-layer threshold logic enhances both sensitivity and specificity while enabling rapid detection of significant anomalies.

5. Experimental Environment, Test Scenarios and Observations

The current section is dedicated to the validation process of the proposed system in an actual home and business setting. The first goal was to evaluate the ability of the developed method to detect ARP spoofing attack within controlled but real-life conditions.

5.1 Setup of the Test Environment

The experimental infrastructure was built using a Raspberry Pi 5 (8 GB RAM) and a Tapo C211 home security camera. The Raspberry Pi established an isolated IoT network and provided internet connectivity exclusively through this segment. IoT devices connected only to the isolated network and had no direct communication with the main home or institutional network. This configuration reflects a common real-world "guest network – IoT network separation" model. The Raspberry Pi was connected to the external network via Ethernet and broadcast a separate wireless network for IoT devices. All ARP traffic within the isolated network was monitored and analyzed in near real-time. During testing, natural traffic was generated through real IoT devices to observe system behavior under typical usage conditions.

5.2 Non-Attack Scenarios (Normal Network Behavior)

In cases of no threat attacks, only the authorized devices within the Internet of Things will be operating in the network. In such circumstances, risk scores have stayed low and stable throughout the experiment. IP-MAC mappings are accurate; ARP packets are at expected levels, while changes in MAC addresses are uncommon. The system allows temporary abnormalities but classifies them as low-risk events; the highest risk score achieved was 12.1 (Figure 3).

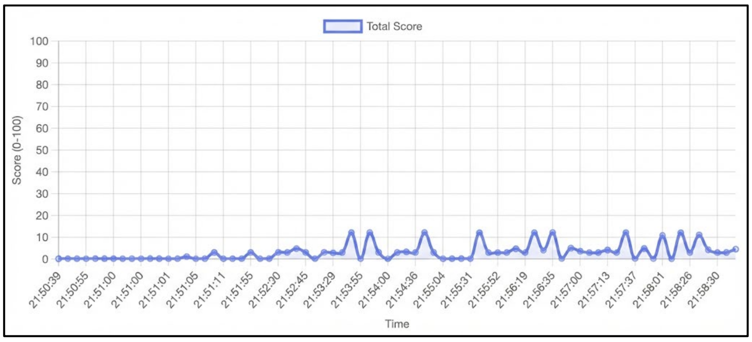

Figure 3. System output total score table in normal network flow

5.3 Attack Scenarios

Controlled ARP spoofing attacks were then executed within the isolated network. The attacker device attempted to manipulate the ARP tables of other devices. Upon initiation of the attack, significant anomalies were detected, particularly in IP–MAC consistency analysis. Rapid changes in IP–MAC mappings and abnormal increases in ARP reply packets led to a rapid escalation in risk scores. As the attack persisted, total risk scores progressively increased, and the system generated "High Risk" and "Attack" level alerts (Figure 4).

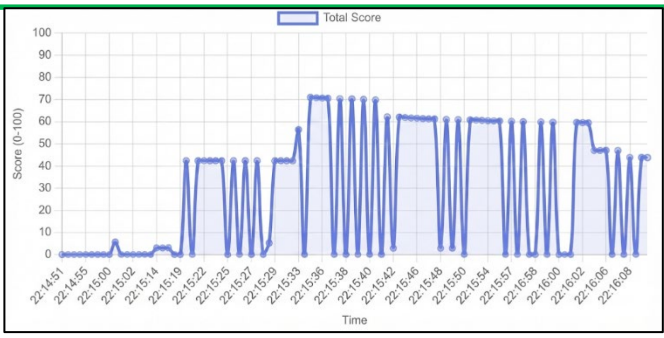

Figure 4. System output of total score table during attack

5.4 Evaluation Criteria and Analysis of Generated Alerts

The evaluation process was concerned with the assessment of criteria based on observation such as alert consistency during attacks, reduction of scores following attacks, and consistency during normal operations. The alerts were set based on accumulated anomalies but not individual packets, thus reducing error due to fluctuations. Moreover, the system responded quickly when high anomalies occurred in each category of analysis. (Figure 5).

Timestamp	IP Address	Total Score	IP-Consistency	Rate Anomaly	Drift Detection	Status
22.12.2025 22:16:35	192.168.4.20	62.43	67.35	59.2	55.0	Attack
22.12.2025 22:16:35	192.168.4.1	0.99	0.0	3.29	0.0	Normal
22.12.2025 22:16:34	192.168.4.20	62.61	67.69	59.2	55.0	Attack
22.12.2025 22:16:34	192.168.4.1	1.14	0.0	3.8	0.0	Normal
22.12.2025 22:16:33	192.168.4.20	62.79	68.05	59.2	55.0	Attack
22.12.2025 22:16:33	192.168.4.1	1.37	0.0	4.56	0.0	Normal
22.12.2025 22:16:32	192.168.4.20	62.97	68.42	59.2	55.0	Attack
22.12.2025 22:16:31	192.168.4.20	63.16	68.79	59.2	55.0	Attack

Figure 5. Attack alerts

5.5 System Performance and Latency

The system operated near real-time on Raspberry Pi 5, with no noticeable latency. Packet capture, analysis, and risk score updates did not disrupt network performance or IoT device functionality. The web-based interface effectively displayed network status, connected devices, and alerts, confirming the system's practicality and usability for home users and small businesses.

6. Discussion and Future Work

The Raspberry Pi-based isolated IoT network architecture proposed in this study provides a low-cost and practical security solution for home and small-scale business environments against ARP spoofing and Man-in-the-Middle attacks. Experimental findings show that the system maintains low risk scores under normal traffic conditions while significantly increasing scores during attack scenarios. By logically separating IoT devices from the main network without requiring VLAN configuration or additional enterprise hardware, the architecture effectively narrows the attack surface. The Raspberry Pi 5 platform successfully performed near real-time ARP analysis without negatively impacting network performance.

The multi-layered detection approach enhances reliability by jointly evaluating IP–MAC consistency, ARP packet rate, and MAC address changes rather than relying on a single metric. Its rule-based design eliminates the need for training data, enabling rapid deployment and adaptability in resource-constrained environments. Although the dynamic learning mechanism reduces false positives, it may slightly increase detection time. The web-based management interface improves usability by presenting visual risk indicators and alerts, making the system accessible to non-expert users. However, the current implementation focuses solely on ARP-layer analysis; DNS spoofing, application-layer threats, and the inspection of encrypted traffic remain outside its scope. Additionally, reliance on a single Raspberry Pi node introduces a potential single point of failure.

In terms of scalability, the proposed system's current architecture targets home and small-scale business environments and can effectively support a limited number of IoT devices. Considering the processing power and memory capacity of the Raspberry Pi 5, an increase in latency and resource usage in the ARP packet analysis process is expected as the number of devices that can be monitored simultaneously increases. Theoretically, it is predicted that the Raspberry Pi 5, with its 8 GB RAM and quad-core processor, can seamlessly support dozens of IoT devices; however, load tests with a large number of devices are required to definitively determine this limit. In this study, only a limited number of real IoT devices could be used in the test environment; scalability evaluation in a large-scale environment could not be performed due to practical limitations. This limitation is considered a research area that is planned to be addressed with more comprehensive testing infrastructures in the future.

In terms of cost comparison, the proposed system offers a significant cost advantage compared to commercial IoT security gateways. Enterprise-class security devices such as Cisco Meraki MX or Fortinet FortiGate require a budget of hundreds to thousands of dollars, including hardware costs and annual license fees. In contrast, the Raspberry Pi 5 hardware used in the proposed system can be obtained for approximately $80-100 USD, with no ongoing license or subscription costs. This cost difference makes the proposed approach an attractive alternative, especially for home users and small businesses that cannot budget for enterprise solutions. While it is acknowledged that the comprehensive feature set and technical support offered by commercial solutions cannot be fully met by this system, the proposed architecture is considered to offer an effective and accessible solution in terms of basic security needs.

Overall, the study demonstrates that low-cost hardware, such as Raspberry Pi, can be effectively used in security-focused network applications. Future work will include automated response mechanisms, hybrid machine learning-based detection models, broader protocol analysis (e.g., DNS, DHCP, MQTT), scalability evaluation in dense IoT environments, distributed multi-node architectures, long-term real-world testing, user behavior-based alerting, and energy efficiency optimization.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. T. R. Hadiningrum, R. A. D. Talasari, K. F. Ilham, and R. M. Ijtihadie, 23, "Survey on Risks Cyber Security in Edge Computing for The Internet of Things Understanding Cyber Attacks Threats and Mitigation", JUTI, vol. 23 no. 1, pp. 29–50, Feb. 2025, doi: 10.12962/j24068535.v23i1.a1210.
2. H. Fereidouni, O. Fadeitcheva, and M. Zalai, "IoT and Man-in-the-Middle Attacks," SECURITY AND PRIVACY, vol. 8, no. 2, Mar. 2025, doi: 10.1002/spy2.70016.
3. K. U. Aditya, P. N. Kamath, Y. Poral, B. D. Mallika and V. Acharya, "Framework for Early Cyber Attack Detection Using ML Models Deployed on Fog Devices," 2024 12th International Symposium on Digital Forensics and Security (ISDFS), San Antonio, TX, USA, 2024, pp. 1-6, doi: 10.1109/ISDFS60797.2024.10527351.
4. Ali, Anas & Husain, Mubashar & Hans, Peter. (2025). Intelligent ARP Spoofing Detection using Multi-layered Machine Learning (ML) Techniques for IoT Networks. 10.48550/arXiv.2507.21087.
5. Imad, Hiba & Abdulridha Hussain, Mohammed. (2022). Defending a wireless LAN against ARP spoofing attacks using a Raspberry Pi. Basrah Researches Sciences. 48. 123-135. 10.56714/bjrs.48.2.12.

6. S. Selvarajan, M. Mohan and B. R. Chandavarkar, "Techniques To Secure Address Resolution Protocol," 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kharagpur, India, 2020, pp. 1-7, doi: 10.1109/ICCCNT49239.2020.9225413.
7. S. Sun, X. Fu, B. Luo and X. Du, "Detecting and Mitigating ARP Attacks in SDN-Based Cloud Environment," IEEE INFOCOM 2020 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), Toronto, ON, Canada, 2020, pp. 659-664, doi: 10.1109/INFOCOMWKSHPS50562.2020.9162965.
8. Karahan, O., & Kaya, B. (2020). Raspberry Pi Firewall and Intrusion Detection System. Journal of Intelligent Systems: Theory and Applications, 3(2), 21-24. https://doi.org/10.38016/jista.653486
9. Adrian, Ronald & Mandasari, R. & Alam, Sahirul. (2025). Design and Implementation of a Machine Learning-Based Adaptive IDS on Raspberry Pi for Smart Home Network Security. Jurnal Sains dan Teknologi Industri. http://dx.doi.org/10.24014/sitekin.v22i2.33485
10. Hussain, Shafiq. (2025). Securing IoT Devices and Edge Computing with Hybrid Mesh Firewalls. https://doi.org/10.13140/RG.2.2.27030.05441.
11. Amin, Rashid & Hussain, Mudassar & Alhameed, Mohammed & Raza, Syed & Jeribi, Fathe & Tahir, Ali. (2020). Edge-Computing with Graph Computation: A Novel Mechanism to Handle Network Intrusion and Address Spoofing in SDN. Computers, Materials & Continua. https://doi.org/10.32604/cmc.2020.011758
12. Girdler, Tom & Vassilakis, Vassilios. (2021). Implementing an intrusion detection and prevention system using Software-Defined Networking: Defending against ARP spoofing attacks and Blacklisted MAC Addresses. Computers & Electrical Engineering. 90. 106990. 10.1016/j.compeleceng.2021.106990.
13. Sharmistha Majumder, Mrinal Kanti Deb Barma, and Ashim Saha. 2024. ARP spoofing detection using machine learning classifiers: an experimental study. Knowl. Inf. Syst. 67, 1 (Jan 2025), 727–766. https://doi.org/10.1007/s10115-024-02219-y

UNEC Journal of Computer Science and Digital Technologies
Volume 2, № 1, 2026, pp. 34-55
https://doi.org/10.30546/UNECCSDT.2026.001.276

Section:
Human–Computer Interaction &
Digital Education Technologies

Barriers to Voice Assistants for Older Adults: Uses and Gratifications Perspective

Ayser Şafak Tuncer[1*], Sezin Eşfer[2]

[1] Faculty of Educational Science, Bahcesehir University, Istanbul, Türkiye*

[1]0009-0008-8490-626X, aysersafak.atuk@bahcesehir.edu.tr

[2]0000-0001-9095-6855, sesfer@gmail.com

Abstract

This study examines how adults aged 60 and over interact with voice assistant technologies, drawing on the Uses and Gratifications (UGT) framework to understand their motivations, expectations, and both the cognitive and communicative barriers they encounter. Employing a qualitative single-case study design, the research involved five older adults (aged 60–75) with diverse demographic and technological backgrounds. Data were collected through semi-structured interviews, scenario-based usability tasks, structured observations, and reflective notes, in line with the depth-oriented goals of qualitative inquiry. Findings indicate that older adults perceive voice assistants as potentially useful tools for managing daily tasks such as reminders, information seeking, and communication; however, actual interactions often increased rather than reduced cognitive load. Despite perceived benefits, adoption was limited by privacy concerns, difficulty formulating commands, distrust of the system's accuracy, and a general sense of technological complexity. Scenario-based tasks increased participants' awareness of the technology's capabilities, yet also surfaced challenges including speech recognition errors, difficulty interpreting system responses, and command formulation problems, all of which heightened cognitive effort and reduced overall satisfaction, particularly among participants with limited prior digital experience. The study highlights the dual nature of voice assistants: they can enhance independence, confidence, and social connectedness, but they may also introduce unexpected demands that complicate the user's experience. These results underscore the importance of designing voice-based Artificial Intelligence (AI) systems that are sensitive to the sensory, cognitive, and communicative needs of older adults. Through the UGT lens, four core gratifications were identified: information access, self-efficacy and personal identity reinforcement, social integration, and entertainment or escape, though the degree to which these were fulfilled varied considerably across scenarios and participants. The study makes a contextual contribution by demonstrating how age-related usability barriers are compounded by language- and culture-specific recognition difficulties in a Turkish-speaking older adult population. Findings offer evidence-based recommendations for the design of age-friendly, linguistically inclusive voice AI systems and for digital inclusion policies targeting older adults.

***Keywords*:** Voice assistants, Uses and Gratifications Theory, Human-AI Interaction, Older Adults

Received:	*Revised:*	*Accepted:*	*Published:*
15/05/2026	*25 /05 /2026*	*05/06/2026*	*17/06/2026*

1. Introduction

As global populations age at an unprecedented pace, the role of digital technologies in supporting the quality of life for older adults has become a central concern in interdisciplinary research. Moving beyond traditional inquiries into user gratifications, contemporary discourse now critically examines why older adults frequently reject voice assistants. Rather than viewing this non-adoption merely as a functional failure or technological aversion, it should be recognized as a deliberate, strategic exercise of agency aimed at maintaining independence and cognitive self-competence [1]. Against the backdrop of escalating care needs, digital health technology adoption has gained considerable momentum, particularly following the shifts in older adults' attitudes and usage behaviours documented in the wake of the COVID-19 pandemic. Recent research demonstrates that voice assistant technologies can serve as effective tools for helping older adults navigate functional and communicative barriers in daily life [2, 3]. These point to the need for user-friendly, age-sensitive solutions that lower barriers to digital participation among older populations.

Voice assistants are AI-driven systems that utilize natural language processing to interpret and execute spoken commands, allowing users to access information, manage tasks, and engage with connected services without the need for physical interaction. By providing hands-free control and intuitive human-computer interaction, these technologies offer a significant approach to addressing the challenges associated with an aging population [45]. Their hands-free and eyes-free interaction modalities make these technologies particularly accessible for older adults who experience difficulties with traditional physical interfaces [3]. Research indicates that usability and emotional needs are key determinants of older adults' attitudes toward voice assistants, with perceived ease of use, security, enjoyment, and sense of companionship found to influence acceptance and usage intentions [4]. Moreover, voice assistants have been shown to yield both psychosocial and functional benefits, including reduction of loneliness, enhancement of social support, and facilitation of safe mobility [5]. However, the extent to which voice assistants improve older adults' quality of life depends on how these technologies are perceived, how they are learned, and the motivations that drive their use [6]. Technology participation extends beyond physical access to encompass cognitive, affective, and communicative dimensions, particularly among populations with functional limitations [7]. Therefore, older adults' interaction processes with voice assistants must be addressed through a multidimensional approach. A recent scholarship on conversational agents and older users further highlights the need to examine dialogue repair strategies, system integration challenges, and emerging health support applications of AI-based voice interfaces [8]. This research investigates the usage of motivations, experiential gratifications, and communicative barriers of individuals aged 60 and over engaging with voice assistant technologies, analyzed within the framework of Uses and Gratifications Theory (UGT).

1.1. Voice Assistant Technologies and Adults Aged 60+

In the contemporary era where technological developments shape human life, the emergence and proliferation of voice assistant technologies has constituted one of the most remarkable phenomena of digital transformation. Prominent applications such as Apple's Siri, Google Assistant, Amazon's Alexa, and Microsoft's Cortana have substantially transformed the nature of human-machine interaction, shifting it from input-based interfaces toward naturalistic spoken dialogue [10]. Unlike the structure of traditional human-computer interaction, which relies on physical and visual interfaces such as keyboards, mice, or touchscreens, voice assistants offer an intuitive communication model based on natural conversational language. This technological revolution creates significant opportunities, particularly for older adults who encounter difficulties in interacting with traditional user interfaces [10]. When considering the barriers that physical and cognitive changes accompanying the aging process create in technology use, the potential offered by voice interfaces becomes even more pronounced [11]. This section addresses the conceptual framework of voice assistant technologies, their historical development process, fundamental technical components, and application areas.

Voice assistants offer meaningful support to older adults in health management through features such as medication reminders, appointment scheduling, and responses to basic health queries, all of which facilitate self-monitoring of health conditions [3, 9]. In addition, voice-activated emergency call features hold particular importance for older adults living alone or managing chronic health conditions [10, 11]. The problem of social isolation, which is prevalent in old age, constitutes one of the important application areas of voice assistants. These technologies can facilitate older adults' communication with family members, provide access to entertainment such as music and news broadcasts, and even offer a degree of social companionship [5, 10].

Voice assistants are used as an important tool in organizing older adults' daily living activities. Functions such as calendar management, creating shopping lists, obtaining weather information, listening to news, and answering various questions support older adults' cognitive capacities [11]. The proactive reminder features offered by this technology are particularly valuable for older adults experiencing memory problems. Furthermore, simplicity of use and adaptability to individual needs are consistently identified as primary expectations among older adults evaluating digital products [4]. The fact that voice assistants are characterized in a manner that can directly meet these demands indicates a theoretically high potential for widespread adoption among the older population. The inadequacy of user manuals and learning difficulties lay the groundwork for the proliferation of voice-guidance-based solutions as tools offering not only entertainment but also education and daily life support [15]. Despite these affordances, significant barriers remain. Speech recognition systems often struggle with older adults' speech patterns, which may include slower speech rates, regional accents, and age-related vocal changes such as reduced volume or altered articulation. System onboarding and initial setup friction also constitute underexamined obstacles to adoption, particularly for individuals with limited digital literacy. Furthermore, emerging LLM-powered voice assistants introduce new possibilities for natural, context-sensitive health dialogue, yet also raise novel concerns regarding information accuracy and over-reliance [1].

1.2. Uses and Gratifications Theory

The origins of Uses and Gratifications Theory (UGT) lie in communication studies, with foundational contributions tracing back to the mid-twentieth century [14]. At the core of the theory are the concepts of "use" and "gratification": use refers to the process by which individuals actively select and engage with media content, while gratification refers to the satisfaction derived from that engagement. Use refers to the process by which individuals select and consume media content, while gratification refers to the sense of satisfaction obtained as a result of this process. The context of this theory concerns why and how individuals consciously choose and use mass communication tools to meet their specific needs [16]. This section focuses on an overview of uses and gratifications theory, its motivational dimensions, and the theory in the context of voice assistant technology for older users.

UGT was among the first theoretical frameworks in communication research to position audience members not as passive recipients of media content but as active agents who deliberately select media to fulfill specific needs and expectations [14, 17]. In this context, individuals are defined not merely as passive receivers of media content, but as "active gratification seekers" who enter into an interaction process with media to meet their expectations and needs. People consider their own motivations and needs in media preferences and make goal-oriented choices accordingly [17].

The concept of gratification extends beyond individual-level psychological satisfaction to encompass broader social processes, including identity formation, normative adaptation, and the negotiation of social roles through media engagement [14, 16]. This concept is not limited to psychological relief or fulfillment of information needs at the individual level; it also encompasses more comprehensive social processes such as identity formation, adaptation to social norms, or developing a stance against these norms. Therefore, the effect of gratification on the individual is not limited to individual experience but can also produce indirect consequences on the social environment and status.The origin of the concept of gratification is based on individuals' tendencies to meet various psychological, cognitive, and social needs during the media use process. Among these needs are different motivations such as access to information, entertainment, socialization, gaining status, and escaping from the pressures of everyday life. These motivations play a guiding role in individuals' selection of specific media content; the gratification obtained from content determines the nature of the relationship established with media [16].

As a result of the media use process, individuals experience certain gratifications. Gratifications sought encompass the benefits and motivations that individuals expect to achieve through media. Active audiences tend to prefer media they believe will meet their expectations, and this is viewed as an important determinant of media usage intention [18]. For example, while one user may select a media tool solely to meet entertainment needs, another may make choices for the purpose of obtaining information or following the agenda. Furthermore, gratifications obtained refer to the benefits that individuals can actually achieve as a result of media use. In this case, the consumer decides whether to use the same media again by considering the gratifications obtained from past experiences [19]. Initial studies on uses and gratifications theory were conducted within the framework of traditional media tools, and over time the research field expanded with the development of the internet and digital technologies. Today, this approach is applied in different contexts ranging from the reasons for using food delivery applications to the adoption of chatbots and online or in-store technologies [17,18, 20].

Applied to older adult populations specifically, research suggests that technology attitudes are closely shaped by individual motivations and social context [5], making UGT a particularly appropriate lens for examining voice assistant adoption in this group. According to UGT, users as active individuals select media content to meet their own needs and obtain different types of gratification as a result of these choices. Particularly when older adults are concerned, the use of voice assistant technologies can provide various gratifications such as obtaining information, entertainment, establishing social interaction, and facilitating daily living activities [12]. In this context, how voice assistant technologies are used by older users and what gratifications this use provides must be comprehensively revealed.

This study addresses a gap in the literature by examining how older adults in Turkey access and experience voice assistant technologies within the context of digital and social participation [2,6]. This study serves as a preliminary study for future longitudinal research on in-home usage, persona comparisons, and experimental interventions based on user interface design. While cognitive, physical, and sensory limitations associated with aging are well-documented barriers to technology use, empirical evidence grounded in older adults' everyday experiences, particularly regarding whether voice assistants effectively reduce these barriers, remains limited [8, 11]. Particularly considering the inadequacy of digital literacy initiatives for older adults in Turkey, revealing how these technologies are experienced in terms of user satisfaction, motivations, and communication barriers encountered provides an original and contextual contribution to digital inclusion discussions. The primary contribution of this study is not positioned within the general voice assistant literature, but rather in its contextual focus on Turkish-speaking older adults. Specifically, the study highlights the intersection of age-related, linguistic, and cultural barriers, making visible how these factors jointly shape user experience [36]. The set of scenarios used in this study is ecologically meaningful for the target audience (medication reminders, making calls, weather/news, music, shopping reminders), and a key strength of this study is that the tasks were selected to be directly related to the daily life practices of older adults, thereby preserving experiential reality. The article's most valuable contribution is contextual; it demonstrates how language- and culture-specific recognition errors exacerbate age-related barriers. Thus, it reveals that voice interfaces, which are assumed to be "universal," can systematically exclude older users who do not speak English. In other words, the study highlights the inequalities that arise at the intersection of aging, language, accent, and digital inclusion [36]. The significance of the research stems from addressing voice assistants not merely as technical innovations but also as communicative and psychosocial tools. In a literature that largely focuses on technical functions, this study proposes an important perspective shift by examining older adults' expectations, gratification levels, and communication barriers within the framework of Uses and Gratifications Theory [14,15]. In this way, the research makes visible potential social benefits such as reducing loneliness, supporting psychological well-being, and enhancing social connectedness, while also aiming to produce evidence-based recommendations for policymakers and technology developers regarding the design of age-friendly digital solutions [5]. This research evaluates the learning processes of individuals aged 60 and over regarding voice assistant technologies, their participation in these technologies, and the communication barriers they encounter within the framework of Uses and Gratifications Theory.

2. Method

This research employed a qualitative case study design, which enables in-depth analysis of a complex phenomenon within its real-life context and is particularly suited to exploratory "how" and "why" questions (29). It documents and interprets the lived experiences of older adults engaging with voice assistant technology for the first time, in line with the experiential case study tradition (29). The phenomenon examined in this research is the interaction of individuals aged 60 and over with voice assistant technologies, their motivations regarding this technology, communicative barriers, and the level of gratification they experience after use.

2.1. Research Design

The research is characterized as a single case study because the situation under investigation encompasses a digital technology experience specific to a particular user group (older adults). This design has made it possible to conduct a holistic analysis without disregarding the cognitive, emotional, social, and environmental contexts of participants regarding technology. Additionally, since the study includes an experiential scenario application, the transformation in participants' perceptions regarding technology use has been monitored comparatively over time.

2.2. Participants

This study involved five (N=5) older adult participants. The participant group consisted of three female and two male older adults, with ages ranging from 60 to 75 years. Educational backgrounds varied between primary school and high school levels, with no participants having university education. The majority of participants (80%) lived with family members, while one participant lived alone (See Table 1). In terms of technology proficiency, most participants (60%) reported basic technology usage levels, while two participants demonstrated intermediate competency. Participants who had some prior experience with voice assistants (P3 and P4) demonstrated relatively higher task performance compared to those with no prior exposure. All participants accessed voice assistants through mobile phone devices, with familiarity spanning different platforms including Siri, Google Assistant, and Alexa. None of the participants reported diagnosed hearing, speech, or visual impairments. However, age-related sensory variation (e.g., reduced vocal volume, slower speech) was observed informally during sessions. All interactions were conducted in Turkish, and Turkish-language Google Assistant was used throughout the study. This diversity in prior exposure, combined with generally limited technology experience, provided valuable insights into the learning curve and usability challenges faced by typical older adult users encountering voice assistant technology. Purposive sampling was employed [21, 22], and the small sample size is consistent with the depth-oriented goals of qualitative case study research. The study does not aim at statistical generalizability; rather, it seeks to generate rich, transferable insights regarding the experiences of a specific population with a bounded technology encounter [29].

Table 1. *Participant Demographics and Technology Usage Background*

Question	P1 (72, Female)	P2 (65, Female)	P3 (60, Female)	P4 (66, Male)	P5 (75, Male)
Gender	Female	Female	Female	Male	Male
Age Range	70–74	65–69	60–64	65–69	75–79
Education Level	High School	Primary School	High School	High School	Primary School
Marital Status	Widowed	Married	Married	Married	Widowed
Living Situation	With Family	With Family	With Family	With Family	Alone
Technology Usage Level	Basic	Basic	Intermediate	Intermediate	Basic
Prior Voice Assistant Use	No	No	Yes	Yes	No
Which Assistant	Siri	Google	Siri	Google	Alexa
Duration of Use	0–3 months	0–3 months	0–3 months	0–3 months	0–3 months
Device	Phone	Phone	Phone	Phone	Phone

The purpose sampling method was used for sample selection [22]. The primary criterion for determining participants was 60 years of age or older and demonstrating basic interest in voice assistant technologies. This sampling approach is appropriate for exploring specific phenomena within particular populations [23].

2.3. Data Collection Tools and Process

Multiple data collection methods were employed in this research, following established protocols for technology usability studies with older adults [24]. The Turkish language was used in all interaction sessions, as Google Assistant was configured in Turkish. This is relevant to interpreting recognition errors, as Turkish-specific names and accents may influence system performance.

Demographic Information Form

A brief demographic form was utilized to determine participants' socio-demographic characteristics. This instrument collected essential background information including age, gender, education level, marital status, living arrangements, prior technology experience, and previous exposure to voice assistant technologies. These demographic variables provided important contextual information for interpreting participants' experiences and identifying potential patterns related to individual characteristics.

Forms

The pre and post experience questions, developed by the researcher under the demographic information forms, were designed to measure changes in participants' perceptions before and after the experience scenarios. The questions assessed variables such as level of knowledge regarding voice assistant capabilities, perceived ease of use, level of confidence in operating the technology, expectations regarding reduced loneliness and socialization opportunities, and intention to continue using voice assistants. The questions were formatted using a 5-point Likert scale. The pre- and post-experience forms primarily served as structured reflective tools to frame participants' awareness before and after the scenarios. Due to the small sample size (N=5), inferential statistical analyses were not performed, and the forms were not treated as psychometrically validated scales. Their function is to provide structured foundations for qualitative interview discussions rather than to produce quantitative outcome measures. Significant changes in individual responses were described descriptively where relevant.

The development of these forms was informed by systematic examination of national and international studies investigating older adults' perceptions of voice assistants [4, 8, 25], thereby ensuring content validity. The literature review encompassed recent empirical research on technology acceptance, usability challenges, and gratifications experienced by older adult users of conversational interfaces. This evidence-based approach to instrument development enhanced the reliability and relevance of the measurement tools for the target population.

Semi-Structured Interview Form

A semi-structured interview guide was developed by the researcher to explore in depth older adults' experiences with voice assistants, including their initial encounters with the technology, purposes of use, perceived benefits and challenges, social and emotional reflections, and intentions for sustainable technology use. The questions were formulated in accordance with the core categories of UGT, specifically addressing information seeking, personal identity, social integration, and entertainment or escape dimensions. The semi-structured format allowed flexibility in exploring emerging themes while maintaining consistency across participants in addressing key research questions.

Experience Scenarios

Brief scenarios were prepared to enable participants to directly experience voice assistant technology in practical contexts. These scenarios included tasks such as medication reminders, calling grandchildren, playing music, and gathering daily information. Each scenario was designed to represent different functional dimensions of voice assistant use, including cognitive support, socialization, sense of independence, and entertainment. The scenarios provided participants with opportunities to engage in natural usage experiences within a controlled environment, facilitating observation of authentic interaction patterns and challenges. This experiential approach allowed participants to form technology perceptions based on actual use rather than abstract speculation.

Observation Notes and Reflective Statements

During the scenario applications, participants' behavioral responses, usage difficulties, and spontaneous verbal expressions were systematically recorded by the researcher using an observation form. This direct observation method captured real-time interactions that might not be fully articulated during subsequent interviews, including non-verbal cues, frustration indicators, expressions of satisfaction, and problem-solving strategies. When deemed appropriate, brief reflective notes were solicited from participants immediately following specific interactions to capture their in-the-moment impressions and emotional responses. These observation data complemented the interview and survey data, providing triangulation that strengthened the overall validity of the findings.

The data collection process consisted of the following phases:

Initial Meeting and Introduction (Maximum 5 minutes): During the initial meeting with participants, the purpose of the research was explained, voice assistant technologies (Siri, Alexa, Google Assistant) were introduced, and preliminary information was gathered regarding participants' demographic information and technology usage habits. This introductory phase is essential for building rapport and establishing baseline technology familiarity [26].

Mini Training Session (Maximum 10 minutes): Participants received brief training on the basic use of Google Assistant. This training demonstrated how to give voice commands, how to activate the device, and basic functions. Brief training sessions are recommended to reduce initial anxiety and provide a foundation for interaction [11].

Scenario-Based Tasks and Individual Experience (Maximum 10 minutes): Participants were asked to complete 5 different scenarios using Google Assistant that reflected situations they might encounter in their daily lives. During this process, participants were encouraged to explore the technology through trial and error. This approach aligns with user-centered design principles that emphasize learning through experience [27].

Semi-Structured Interview (Maximum 20 minutes): Following completion of the scenario tasks, semi-structured interviews were conducted with participants. During these interviews, questions were posed regarding their experiences, challenges they encountered, attitudes toward technology, and potential for daily life usage. Semi-structured interviews allow for flexibility while maintaining focus on key research questions [21].

Final Assessment Questions: Semi-structured interview questions were administered to assess participants' interaction skills with the voice assistant, followed by final assessment questions. Test results were graded out of 65 points, and percentages were calculated. Usability testing provides quantitative measures to complement qualitative insights [28]. The authors acknowledge that the short training and scenario exposure (approximately 20 minutes total) limited the ecological validity of the findings and hindered the assessment of learning curves or sustained use. Therefore, the study was positioned as an initial, exploratory encounter study rather than a longitudinal or experimental investigation [29].

Usability Assessment

A structured usability observation rubric was used to evaluate participants' task completion during the five scenarios. The total possible score was 65 points, distributed as follows: each of the five scenarios was assessed on a 13-point scale covering (a) task initiation (0–3), (b) command formulation accuracy (0–3), (c) response comprehension (0–3), (d) error recovery (0–2), and (e) task completion (0–2). Scoring was conducted by the first author, who directly observed each session. Due to the single-evaluator setup, inter-rater reliability for the rubric itself was not established. Scoring was conducted independently by the first and second authors using the same usability rubric. To ensure inter-rater reliability, agreement between the two evaluators was calculated using the formula proposed by Miles and Huberman, resulting in an agreement rate of 85% [39]. Discrepancies in scoring were discussed and resolved through consensus, and the final scores were determined accordingly. The authors note this as a limitation and recommend that future studies employ dual-rater scoring with reported inter-rater reliability coefficients. The rubric scores should therefore be interpreted as indicative rather than definitive measures of task performance.

Scenarios

The scenarios used in the research were designed considering the needs that older adults might encounter in their daily lives.

Health Tracking - Medication Reminder Scenario: "Remind me to take my blood pressure medication every morning at 9:00 a.m." This scenario aims to evaluate participants' ability to use the voice assistant for health management, which is a critical need for older adults managing chronic conditions.

Social Interaction - Communicating with Grandchildren Scenario: "Hey Google, call Arda." This scenario examines the potential of voice assistants to facilitate social connections, which are vital for older adults' well-being.

Information Gathering and Staying Up-to-Date Scenario: "What's the weather like in Istanbul today?" and "Read today's news." This scenario evaluates participants' experiences in meeting daily information needs, supporting cognitive engagement, and autonomy.

Entertainment and Enjoyment - Music Playing Scenario: "Play a song by Müzeyyen Senar." This scenario aims to explore how technology can be used in leisure activities, which contribute to quality of life.

Cognitive Support and Daily Organization - Reminder List Scenario: "Remind me to buy milk and bread tomorrow morning." This scenario examines the support provided by voice assistants in daily life organizations, addressing common memory concerns among older adults.

In scenarios involving contact names (Scenario 2) and culturally specific artist names (Scenario 4), the researcher observed and noted instances where recognition errors occurred. These errors were associated with Turkish phonology and naming conventions that may not be well-represented in the assistant's default language model. This contextual factor is discussed as a source of communicative friction in the results section. The study did not experimentally manipulate phrasing alternatives; however, in cases of repeated recognition failure, the researcher offered a rephrased prompt as a naturalistic support measure.

2.4. Data Analysis

Two primary analysis methods were employed in the research. First, participants' processes of completing scenario tasks were recorded through observation, and interaction behaviors and emotional responses were noted. Observation is a fundamental method for understanding real-time user interactions with technology [29]. Second, using contextual analysis methods, participants' experiences, technology usage environments, and needs were evaluated from a holistic perspective. Data obtained from semi-structured interviews were analyzed through content analysis, and recurring themes and patterns were identified [30]. This thematic approach allows for systematic identification of meaningful patterns across qualitative data. Theme development followed an iterative process; in the first stage, the first author conducted open coding by examining the interview transcripts and observation notes line-by-line, generating initial codes grounded in participants' expressions and interaction experiences.

In the second stage, these initial codes were reviewed, compared, and grouped into broader categories. Through an inductive process, a preliminary codebook was developed, including code definitions, inclusion criteria, and representative examples. This codebook serves as a guiding framework for organizing and refining themes. The coding structure and emerging themes were then reviewed and refined in collaboration with the second author to ensure conceptual clarity and consistency. During this stage, codes were merged, revised, or redefined where necessary. To ensure coding reliability, a subset of the data was independently coded by the second author using the established codebook. For the usability rubric, inter-rater agreement between the two authors was calculated at 85% using the formula proposed by Miles and Huberman [39]. A separate inter-coder reliability analysis was conducted for thematic coding, also yielding an 85% agreement rate using the same formula [39]. Discrepancies between coders were discussed and resolved through consensus, and the codebook was finalized accordingly. Thus, the coding process was conducted through a collaborative and iterative procedure between the first and second authors, ensuring both analytical rigor and consistency. The codebook was developed inductively, with themes grounded in participants' own language and experiences. Given the small sample (N=5), formal data saturation was not assessed; instead, thematic consistency across participants served as an internal criterion for evaluating theme robustness, consistent with qualitative case study practice [29, 30]. Representative quotations are provided for each theme to support transparency.

In qualitative research, validity is primarily considered within the framework of internal and external validity dimensions. Internal validity is related to the extent to which the research accurately, deeply, and holistically reflects the phenomenon it addresses. In this context, it is important for the researcher to adopt a consistent, systematic, and rigorous approach in the data collection, analysis, and interpretation phases [30]. In this study, in order to increase internal validity, the findings obtained were presented with detailed descriptions; the data were first presented at a descriptive level and in an unbiased manner, and then the analytical and interpretive analysis process was initiated. Furthermore, the principles of homogeneity and heterogeneity were considered to clearly reveal the commonalities and differing characteristics among the data. External validity, on the other hand, is concerned with the applicability and transferability of the research results in similar contexts [21,30]. Accordingly, the data analysis process was explained in detail, aiming to make the research traceable and replicable by other studies.

Reliability in qualitative research is directly related to the transparency and traceability of the process and the consistency of the findings; in other words, it expresses the degree to which similar results can be obtained when using the same methods [30]. In this study, to ensure reliability, the data analysis process was carried out by two independent experts, and the resulting coding results were compared. The level of agreement between the coders was calculated using the reliability coefficient formula proposed by Miles and Huberman (1994); the reliability rate was determined to be 85%. This value indicates that the research findings are reliable and consistent.

2.5. Ethical Considerations

Prior to the research, participants were provided with detailed information about the purpose of the study, the process, and data usage, and their voluntary participation consent was obtained. Participants' personal information was kept confidential, and anonymity was ensured using a coding system (P1, P2, P3, P4, P5).

3. Results

The usability test results across all five scenarios revealed varying levels of success among participants in completing voice assistant tasks. Table 2 presents the overall performance scores for each participant. Scores reflect task performance as observed and scored by the researcher using the structured rubric described in Section 2.4. These scores are intended as descriptive indicators of performance, not as psychometrically validated measures.

Table 2. *Participant Demographics and Usability Test Scores*

Participant	Age	Gender	Score	Max Score	Percentage
P1	72	Female	38	65	58.46%
P2	65	Female	38	65	58.46%
P3	60	Female	43	65	66.15%
P4	66	Male	36	65	55.38%
P5	75	Male	39	65	60.00%

The results indicate that participant performance ranged from 55.38% to 66.15%, with P3 achieving the highest score and P4 the lowest. The average success rate across all participants was approximately 59.69%, suggesting moderate overall usability but also highlighting significant room for improvement in voice assistant design for older adult users. In addition to overall usability scores, scenario-based performance patterns were examined descriptively to enhance analytical transparency. Descriptive reporting is particularly valuable in small-sample qualitative studies, as it supports traceability and provides a clearer understanding of observed interaction patterns [29, 39].

Participants demonstrated relatively higher success in simpler, single-step tasks such as information retrieval and music playback. In contrast, lower performance was observed in scenarios requiring multi-step command formulation, such as setting reminders and initiating contact-based communication. Recognition-related difficulties were more frequently observed in scenarios involving proper names (e.g., calling a contact or requesting music by artist name), whereas comprehension-related challenges emerged in scenarios where the system produced longer or more complex responses. These descriptive patterns provide additional insight into where interaction breakdowns most commonly occur across scenarios.

Table 3. *Scenario-Based Performance and Observed Difficulties*

Scenario	Task Description	Performance Level	Main Observed Difficulties
Scenario 1	Medication Reminder - Health Tracking	Moderate	Confusion between alarm and reminder Difficulty in command repetition
Scenario 2	Communicating with Grandchildren - Social Interaction	Low–Moderate	Name recognition errors (e.g., “Arda”) Uncertainty about system response

Scenario 3	Daily Information Gathering - Staying Informed	High	Long and complex responses Difficulty controlling information flow
Scenario 4	Entertainment and Enjoyment - Music Playing	High	Artist name recognition errors Limited control over playback
Scenario 5	Reminder List - Cognitive Support and Daily Organization	Moderate	Difficulty with multi-step commands Uncertainty in managing saved reminders

Scenario 1: Medication Reminder - Health Tracking

The medication reminder scenario, which required participants to command "Remind me to take my blood pressure medication every morning at 9:00 a.m.," revealed significant insights into how older adults perceive and interact with voice assistant technology in health management contexts.

Positive Outcomes

Participants demonstrated three primary positive experiences when successfully completing this health-related task. First, participants expressed a notable sense of independence, indicating that the ability to manage their medication schedules autonomously through voice commands contributed to their self-reliance. Second, the medication reminder function provided participants with a sense of security, as automated reminders reduced anxiety about forgetting critical health-related tasks. Third, participants perceived that the voice assistant gave them greater control over their health management.

Challenges and Negative Experiences

Despite these positive outcomes, participants encountered notable difficulties that hindered the full effectiveness of the medication reminder scenario. A primary challenge was confusion between alarms and reminders. Several participants expressed confusion between alarm and reminder functions, indicating difficulty in distinguishing between these two system features during task completion. Additionally, several participants reported forgetting to repeat commands when the voice assistant failed to recognize their initial input. This pattern indicates potential issues with both the technology’s speech recognition accuracy for older adults’ voices and participants’ understanding of error recovery procedures. This pattern indicates potential issues with both the technology's speech recognition accuracy for older adults' voices and participants' understanding of error recovery procedures. The need to repeat commands can create frustration and reduce perceived usability, particularly when immediate success is not achieved. This was particularly evident with Turkish given names (e.g., “Arda”), where Turkish phonological patterns including vowel harmony and specific consonant clusters may not align well with the default recognition weightings in standard Google Assistant models. Recognition errors of this nature have been documented in multilingual and non-English voice assistant deployments [36], and represent a systemic equity concern for non-anglophone older user populations.

Scenario 2: Communicating with Grandchildren - Social Interaction

The social interaction scenario required participants to use the voice command "Hey Google, call Arda" to initiate a phone call with a family member, specifically designed to assess the technology's role in maintaining intergenerational connections.

Positive Outcomes

Participants exhibited strong positive emotional responses when successfully completing this task. The ease of initiating calls through voice commands eliminated several barriers that traditionally complicate phone usage for older adults, such as navigating contact lists, remembering phone numbers, or managing touchscreen interfaces. The scenario also revealed that voice-activated calling reduced the cognitive load associated with traditional phone usage.

Challenges and Negative Experiences

Several participants struggled with the proper pronunciation and enunciation required for the voice assistant to recognize contact names accurately. This was particularly evident with Turkish given names (e.g., “Arda”), where Turkish phonological patterns — including vowel harmony and specific consonant clusters — may not align well with the default recognition weightings in standard Google Assistant models. Recognition errors of this nature have been documented in multilingual and non-English voice assistant deployments [36] and represent a systemic equity concern for non-anglophone older user populations. Additionally, some participants expressed uncertainty about whether the command had been successfully executed, as the feedback provided by the voice assistant was not always immediately clear or sufficiently prominent.

Scenario 3: Daily Information Gathering - Staying Informed

This scenario involved two related tasks: asking for weather information ("What's the weather like in Istanbul today?") and requesting news updates ("Read today's news"). These tasks assessed participants' ability to use voice assistants for accessing timely, practical information.

Positive Outcomes

Participants generally found the information retrieval of tasks to be straightforward and valuable. The immediate access to weather information was particularly appreciated, as it directly supported daily decision-making regarding activities and appropriate clothing. The conversational nature of the queries aligned well with participants' natural speech patterns, making the interaction feel intuitive. The news reading function was also well-received by participants who expressed interest in staying informed about current events. For those with visual impairments or reading difficulties, the audio format provided an accessible alternative to traditional news consumption methods. Participants noted that this feature could potentially reduce their reliance on family members to relay information, thereby enhancing independence.

Challenges and Negative Experiences

Despite these benefits, participants encountered several difficulties. The voice assistant's responses to weather queries were sometimes overly detailed or technical, with confusing or unnecessary terminology. Participants expressed a preference for concise, actionable information rather than comprehensive meteorological data. The news reading function presented additional challenges. Participants found it difficult to control the pace and selection of news stories, with some expressing frustration at being unable to easily skip topics of interest or replay information they had missed. The linear, non-interactive nature of the news delivery contrasted with participants' expectations of being able to navigate content more flexibly, as they might with printed newspapers or television news programs.

Scenario 4: Entertainment and Enjoyment - Music Playing

The entertainment scenario asked participants to request music by commanding "Play a song by Müzeyyen Senar," assessing the voice assistant's role in leisure activities and cultural engagement.

Positive Outcomes

This scenario elicited some of the most positive emotional responses from participants. The ability to instantly access music from preferred artists through simple voice commands was highly valued, particularly for participants with mobility limitations or difficulties operating traditional media devices. Participants expressed delight at the ease with which they could enjoy music that held personal or cultural significance. The music playing function also demonstrated potential for mood regulation and emotional well-being. Participants noted that having immediate access to preferred music could help alleviate feelings of loneliness or boredom, providing a form of companionship and comfort. The nostalgic value of accessing familiar artists and songs appeared to contribute significantly to participants' positive perceptions of this feature.

Challenges and Negative Experiences

Some participants struggled with the pronunciation of artist names. Specifically, the name “Müzeyyen Senar” (a culturally significant Turkish musical artist) was frequently misrecognized, either resulting in no match or an incorrect artist. This finding illustrates how culturally non-dominant names function as points of failure in voice assistant systems that are trained predominantly on anglophone or high-resource-language data [36]. Additionally, participants expressed limited understanding of how to control playback once music began playing, with commands for pausing, skipping, or adjusting volume not being intuitive to all participants.

Scenario 5: Reminder List - Cognitive Support and Daily Organization

The final scenario required participants to set up a shopping reminder by commanding "Remind me to buy milk and bread tomorrow morning," testing the voice assistant's utility for everyday organizational tasks and cognitive support.

Positive Outcomes

Participants recognized the practical value of this feature for managing daily tasks and compensating memory concerns. The ability to quickly capture reminders without needing to write them down or navigate complex applications was seen as highly beneficial. Participants appreciated that the voice assistant could serve as an external memory aid, reducing the cognitive burden of remembering multiple tasks throughout the day. The natural language processing capabilities of the assistant were particularly evident in this scenario, as participants could specify items and timing in conversational language rather than following rigid command structures. This flexibility made the interaction feel more natural and less like operating a traditional computing device.

Challenges and Negative Experiences

Despite these advantages, participants encountered difficulties with reminder management. Once reminders were set, participants were uncertain about how to review, modify, or delete them using voice commands alone. This limitation meant that participants could not easily verify that their reminders had been correctly understood and stored by the system. Additionally, some participants struggled with formulating commands that included multiple pieces of information, such as both the items to remember and the timing of the reminder. When commands were complex or included in multiple items in a list, the voice assistant sometimes failed to capture all the information accurately, requiring participants to repeat or rephrase their requests.

Speech Recognition Errors and User Coping Strategies

Across all scenarios, speech recognition errors emerged as a recurring interaction barrier. These errors were particularly evident in tasks involving proper names and culturally specific content, such as contact names (e.g., “Arda”) and artist names (e.g., “Müzeyyen Senar”). At the participant level, recognition-related errors occurred at least once per participant in scenarios involving name-based commands, particularly in Scenario 2 (calling a contact) and Scenario 4 (music playback). While the small sample size does not allow for statistical generalization, this pattern was consistently observed across participants. The types of recognition problems included misinterpretation of names, failure to detect the intended entity, and incorrect substitutions with phonetically similar words. These issues were more pronounced when participants spoke with lower vocal intensity or slower articulation.

Participants employed several coping strategies when encountering recognition errors. The most common strategy was repeating the same command without modification. Some participants attempted to increase vocal clarity by speaking louder or more slowly. A smaller number of participants used reformulation strategies, such as adding contextual cues (e.g., “my grandson Arda”), which were more effective in achieving successful outcomes. However, most participants did not spontaneously adopt adaptive reformulation strategies, indicating limited awareness of effective interaction repair mechanisms. This resulted in increased cognitive load and occasional frustration during task completion. These findings highlight the need for voice assistant systems to provide more explicit feedback and guided error-recovery support, particularly for older users interacting in non-English contexts.

Table 4. *Thematic Analysis Results*

Theme	Sub themes	Codes	Example Quotes
Theme 1: Cognitive Overload and Comprehension Barriers	Attention and Information Overload Error-Induced Cognitive Disruption	• Difficulty formulating commands • Loss of attention with long information • Cognitive disruption/panic caused by incorrect responses • Increased cognitive load due to follow-up questions	"I got confused; I didn't know what to say." (P1) "I got bored when the news was too long." (P3) "It said something else, I panicked." (P2) "It asked 'What time?' What's the point?" (P5)
Theme 2: Technical Barriers and Loss of Control	Perceived Complexity and Error Anxiety Interface and Navigation Confusion Low Perceived Autonomy	• Fear of making mistakes • Perceiving technology as complicated • Confusion about apps/subscriptions • Follow-up questions perceived as unnecessary • Lack of perceived independence	"I'm afraid it will call the wrong person." (P2) "It's too complicated; I can't grasp it." (P5) "When Spotify Premium popped up, it ruined my mood." (P4) "It asked 'What time? 'what's the point?" (P5) "I didn't think of it that way; independence didn't occur to me." (P1)
Theme 3: Social-Emotional Deficits	Emotionally Meaningful Functional Use Absence of Social Companionship	• Not perceiving the assistant as a social partner • Emotional connection formed primarily through music •Preference for human interaction over AI assistant	"No, it doesn't reduce loneliness." (P2) "When it played Müzeyyen Senar, it lifted my mood." (P4)

The thematic analysis revealed three interrelated patterns that characterize older adults' experiences with voice assistants. First, participants reported significant cognitive challenges during interaction, including difficulty formulating commands, sustaining attention when information was lengthy, and managing follow-up questions. Incorrect or unexpected responses often led to confusion or anxiety, indicating that such interactions increased cognitive load and disrupted information processing rather than facilitating it. Second, the findings point to perceptions of technical complexity and a diminished sense of control, which functioned as barriers to independent use. Participants frequently expressed fear of making mistakes, viewed the technology as overly complicated, and experienced confusion regarding applications, subscriptions, or system prompts. Follow-up questions were commonly perceived as unnecessary and intrusive, reinforcing feelings of dependence rather than autonomy. Finally, the analysis showed limited social and emotional gratification associated with voice assistant use. Most participants did not conceptualize the assistant as a social or relational partner, instead positioning it primarily as a functional tool. Emotional engagement was largely confined to specific features, particularly music playback, which occasionally enhanced mood but did not substantially alleviate feelings of loneliness or foster a sustained sense of social connection. Three participants (P1, P2, P5) spontaneously expressed discomfort regarding potential data retention and the sense of being continuously monitored. Among participants living with family members (P1 and P2), concerns centered primarily on privacy from external parties, whereas P5, who lived alone, expressed concerns more closely related to perceived surveillance by family members who might access the device. Concerns of this type have been documented in the literature on smart speaker adoption and are particularly salient for older adults with limited prior exposure to always-on devices [37].

4. Discussion

The findings from this case study reveal that voice assistants partially compensate for cognitive and physical limitations commonly experienced by older adults, providing practicality, time savings, and a sense of independence in daily activities [31]. These outcomes align with broader research on assistive technologies emphasizing the importance of supporting autonomy and reducing functional barriers in aging populations [6, 11]. Analysis through the UGT framework revealed four core gratifications that emerged from participants' interactions with voice assistants: information seeking, self-efficacy and personal identity reinforcement, social integration, and entertainment or escape [16]. These gratifications represent the underlying motivations and satisfactions that drive older adults' adoption and continued use of voice assistant technology. Participants reported increased confidence and a greater sense of control following the medication reminder and shopping list scenarios, suggesting that voice-activated reminder functions may support self-management and reduce cognitive burden among older adults managing daily health routines [9]. The ability to externalize memory tasks through simple voice commands appeared to provide psychological relief and reinforce participants' sense of competence.

Item-level shifts observed in pre- and post-experience reflections can be interpreted through the UGT framework in terms of the gap between gratifications sought and gratifications obtained [14, 18]. Participants initially expressed expectations related to ease of use, independence, and social connection (gratifications sought). However, the actual interaction experience revealed a more complex outcome, where certain gratifications were achieved (e.g., entertainment and basic information access), while others were only partially fulfilled or remained unmet due to interaction difficulties (gratifications obtained). In particular, while participants anticipated that voice assistants would reduce cognitive effort and support autonomy, the need to repeat commands, manage errors, and interpret system responses increased cognitive load in some scenarios. This discrepancy highlights a gap between expected and experienced gratifications, suggesting that usability challenges directly influence the continuity of technology use among older adults [38].

The communication scenario, particularly the task of calling grandchildren, supported the maintenance of social bonds and intergenerational connections. Voice assistants facilitated emotional closeness and social connectedness by reducing technical barriers to communication [10]. This finding is particularly significant given the well-documented relationship between social isolation and negative health outcomes among older adults. The technology's potential to strengthen family ties and reduce feelings of loneliness emerged as a critical benefit that extends beyond mere functional utility. The music playback scenario produced the strongest emotional satisfaction among all tested tasks. The nostalgic value of accessing culturally significant music triggered positive emotional responses and strengthened participants' attitudes toward technology [32]. This suggests that voice assistants may serve an important role in supporting emotional well-being and maintaining cultural identity among older adults, particularly when the technology can accommodate diverse musical preferences and artist recognition. Despite these benefits, several significant barriers emerged that hindered optimal use and acceptance of voice assistant technology. Complexity and cognitive load represented key obstacles to successful interaction. Long audio responses, conceptual confusion between similar functions (such as alarms versus reminders), and premium application requirements reduced participants' motivation to continue using the technology [23,33]. These findings underscore the importance of designing voice interfaces that align with older adults' cognitive processing capabilities and expectations.

Participants consistently expressed a preference for short, clear, single-step commands with immediate and predictable outcomes, reflecting established principles of age-friendly interface design emphasizing simplicity, error tolerance, and transparency [6, 11]. The confusion between alarms and reminders observed in this study reflects a broader design challenge: voice assistant architectures often distinguish between these functions in ways that are not transparent or intuitive to users whose mental models are shaped by non-digital practices [35]. Designers are encouraged to explore disambiguation prompts that are brief, confirmatory in tone, and do not require users to recall technical category names (e.g., “Should I set this as a recurring reminder?” rather than “Would you like an alarm or a reminder?”).

The recognition errors associated with Turkish proper names and artist names in this study highlight a systemic issue in voice assistant accessibility: the technology’s performance is not language- or culture-neutral. Older adults who interact in less-resourced languages or with culturally specific vocabularies encounter a higher frequency of interaction failures that are not attributable to user error. This finding is consistent with Seaborn et al. [36], who documented similar challenges in cross-cultural voice assistant deployments, and underscores the importance of culturally localized training data and fallback dialogue strategies in assistive voice technologies.

4.1 Design and Implementation Recommendations

The findings of this study provide several data-driven design implications for improving voice assistant interactions for older adults. First, in critical tasks such as setting reminders or initiating communication, systems should explicitly confirm the recognized command and clearly state the intended action. Rephrasing the user’s input (e.g., “I will call Arda now”) can enhance user trust, reduce uncertainty, and increase error awareness. This type of confirmatory feedback is particularly important for older users who may experience elevated cognitive load and uncertainty during novel technology interactions [9, 43].

Second, interaction design should support guided error recovery. Rather than requiring users to independently reformulate commands, systems should provide simple and adaptive prompts (e.g., “Did you mean Arda from your contacts?”). Such support mechanisms can reduce frustration and facilitate more effective interaction repair, especially for users with limited digital literacy.

Third, localization should be considered beyond direct language translation. Voice assistant systems should incorporate culturally specific elements such as personal names, media preferences, and everyday communication practices. The findings indicate that recognition failures were more frequent in culturally specific inputs, suggesting that language models should be adapted to local linguistic and cultural contexts. This aligns with research emphasizing the importance of culturally sensitive AI systems [36].

Fourth, voice interfaces should be designed to minimize cognitive load by providing shorter, clearer responses and avoiding unnecessary follow-up questions. Older adults demonstrated a clear preference for simple, predictable, and single-step interactions. Designing for reduced cognitive demand is a key principle in age-friendly technology development [6]. In addition, several interaction design principles can be derived from the observed user behaviors.

Longer silence tolerance during voice input, clear listening indicators, and concise system responses can support older users' sense of control and comprehension. Participants demonstrated difficulty following extended responses and managing interaction flow, suggesting that shorter summaries and more structured dialogue transitions may improve usability. Reliable topic transitions and clear feedback mechanisms can further reduce uncertainty and enhance user confidence during interaction. These principles are consistent with prior research emphasizing the importance of cognitive load reduction and clear interaction cues in age-friendly interface design [6].

Finally, inclusive speech recognition should be prioritized. Systems should better accommodate variations in accent, pronunciation, and speech pace associated with aging. These findings suggest that recognition limitations are not solely technical issues but also relate to accessibility and digital inclusion. Designing more inclusive voice interaction systems can help reduce disparities in technology use among older populations [36].

5. Conclusion and Recommendations

This research demonstrates that voice assistants can provide meaningful benefits in terms of practicality, independence, and confidence for individuals aged 60 and older. Through the UGT framework, four gratification dimensions were identified as relevant to participants' voice assistant interactions: information access, self-efficacy and personal identity reinforcement, social integration, and entertainment or escape. However, the degree to which these gratifications were fulfilled varied considerably across tasks and individuals. The music playback and communication functions produced the highest emotional impact, suggesting that voice assistants hold particular promise for supporting emotional well-being and social connection among older adults. However, significant usability challenges persist. Conceptual confusion between similar functions, excessively long or complex responses, and premium service barriers emerged as the primary obstacles to adoption and sustained use. Older adults demonstrated clear preferences for short, clear, single-step commands that minimize cognitive load and produce predictable outcomes. The findings suggest that accent- and name-recognition failures should not be framed solely as technical limitations but also as matters of linguistic justice and digital equity, particularly for non-anglophone older users [36, 40]. Voice assistant systems that fail to adequately recognize culturally and linguistically diverse inputs may unintentionally exclude older users, particularly those interacting in non-English contexts. Addressing these challenges is therefore essential not only for improving usability but also for promoting equitable access to digital technologies.

Future research should prioritize the development of validated measurement scales specifically designed to assess age-friendly voice assistant features, evaluating command simplicity, feedback clarity, error tolerance, and emotional engagement. Longitudinal research is essential to examine habit formation, sustained motivation, and technology abandonment. The social context of technology use, particularly family involvement, deserves greater attention. Structured voice assistant literacy programs for older adults should emphasize short commands, error-tolerant strategies, and visual guidance. Technical improvements to voice assistant design should include summary modes for concise responses, culturally sensitive name matching algorithms accommodating Turkish and other non-anglophone naming conventions, simplified onboarding procedures, and comparative platform evaluations that explicitly address accessibility dimensions for older users [35-36].

The findings should be interpreted specifically within the context of Google Assistant used via smartphone, as older adults' experiences with voice assistants may vary across platforms, devices, and interaction modalities [42]. In addition, this study examined initial experiences in controlled scenarios rather than prolonged use in natural home settings; however, prior longitudinal research suggests that older adults' perceptions and usage patterns may change over time in real-world contexts [41]. Cognitive load in the present study was interpreted primarily through observational and self-reported data, and future research could strengthen this interpretation by incorporating more objective indicators such as reformulation frequency, task completion time, and scenario-based error rates. Support mechanisms should also be designed to guide older adults without undermining user autonomy, particularly during early use, when simple and repeatable learning supports may be especially beneficial [43]. Finally, future studies should broaden their ethical considerations by explicitly addressing data protection, boundaries of use, and safeguards against over-reliance on technology.

In addition to the main findings, several nuanced observations emerged regarding prior experience, system response design, and privacy perceptions. Participants with limited prior experience using voice assistants (P3 and P4) demonstrated relatively higher task performance compared to others. Their familiarity with basic command structures and interaction flow enabled them to complete tasks more fluently. However, prior experience did not eliminate interaction challenges. These participants also encountered difficulties related to name recognition, command formulation, and understanding system responses. Although they appeared more confident and slightly less cognitively strained, confusion still occurred, particularly in multi-step tasks and situations involving a long or ambiguous system of feedback. This suggests that prior exposure facilitates interaction but does not fully resolve usability challenges for older adults.

Another important finding concerns the design of system responses. Participants had difficulty following long and information-dense responses. In particular, responses that included multiple pieces of information delivered consecutively—such as weather reports containing temperature, humidity, and wind details—were perceived as overwhelming. Similarly, continuous delivery of news content led to loss of attention and reduced comprehension. Participants consistently expressed a preference for shorter, simpler, and more direct responses. These findings highlight the importance of managing information density and response length in voice interaction design for older users.

Finally, privacy perceptions emerged as an influential factor in user experience. Although participants were provided with a general explanation at the beginning of the study regarding how the system operates, some participants still expressed concerns about being monitored during interaction. This indicates that initial explanations alone may not be sufficient to establish trust. Repeated and clearer privacy-related explanations throughout the interaction process could potentially enhance user confidence and acceptance. However, it should also be noted that deeply rooted beliefs and concerns about technology and data security among older adults may not be easily mitigated through information alone. Trust appears to develop gradually over time through repeated and positive interaction experiences rather than through one-time explanations. In conclusion, voice assistants have substantial potential to improve quality of life for older adults, but realizing this potential requires age-friendly, accessible, and cognitively appropriate design. The technology must be developed with explicit consideration of the unique needs, capabilities, and preferences of older adult users, rather than simply adapting designs created for younger, more technologically proficient populations.

6. Limitations

The study’s use of Google Assistant exclusively is acknowledged as a limitation of generalizability. Comparative platform studies for instance, examining touch-enabled Alexa Echo Show against voice-only interfaces could clarify how multimodal feedback affects usability for older adults and helps designers understand trade-offs between voice, visual, and tactile interaction modalities [37]. The brief, single-session exposure used in this study is also a significant limitation: it precludes any assessment of habituation, learning curves, or evolving attitudes over time. Longitudinal or diary-based home-use studies would substantially strengthen the evidence base for this research area.

The study is explicitly positioned as a bounded, exploratory case study with five participants, conducted within a single structured session. The findings are not intended to generalize to the broader older adult population but rather to contribute transferable insights and to identify design and research priorities. The following limitations should be noted: (1) the small sample (N=5) limits statistical and broad empirical generalizability; (2) the brief session duration limits the capture of learning effects or habituation; (3) the single-platform (Google Assistant, Turkish language) design limits cross-platform comparability; and (4) the pre-/post-test forms, while useful as reflective anchors, were not treated as validated instruments and yielded no inferential data.

This study should be interpreted as a case study rather than a basis for broad generalization. The very small sample size (N = 5), the short single-session exposure, and the guided nature of the scenarios limit transferability, ecological validity, and any meaningful assessment of learning effects over time. This limitation is especially important because prior longitudinal research has shown that older adults' perceptions and uses of voice assistants evolve with sustained real-world experience, and that initial convenience or novelty may not persist in longer-term use [41]. In addition, the study was restricted to a single platform (Google Assistant) and a single interaction modality (phone-based voice use), which narrows the generalizability of findings across devices, ecosystems, and multimodal interfaces. This is relevant because older adults' intentions to use voice assistants are shaped by multiple factors beyond usability, including emotional needs, privacy concerns, self-efficacy, and perceived companionship [42]. Methodologically, although the structured usability rubric improves analytic transparency, the use of a single evaluator limits inter-rater reliability, and the qualitative coding procedure would benefit from more explicit reporting of coder roles, codebook development, and agreement calculation. The absence of systematic measurement of speech-recognition failures further constrains interpretation, particularly because recent ASR scholarship has shown that accent-related recognition problems remain conceptually and technically under-addressed, with implications for accessibility and linguistic equity [40]. Nevertheless, the study makes a meaningful contextual contribution by showing how age-related barriers may be compounded by language- and culture-specific recognition difficulties. Prior work has also demonstrated that communication style and anthropomorphic framing can significantly influence older adults' trust, acceptance, and mental workload when interacting with voice assistants, suggesting that future studies should move toward larger and more diverse samples, in-home longitudinal designs, and experimental comparisons of conversational style, disambiguation strategies, feedback design, and localized ASR adaptations for older users [43].

Authors' Declaration

The authors declare that there are no conflicts of interest related to this study.

Authors' Contribution Statement

The first author conceptualized the research topic as part of their master's thesis, conducted the data collection process, performed data analysis, and prepared the initial draft of the manuscript. The second author, serving as the thesis supervisor, provided guidance throughout all research phases, contributed to the research design, supervised the data analysis process, and reviewed and edited the manuscript for publication.

Author Notes

Based on *Academic Integrity and Transparency in AI-assisted Research and Specification Framework* (44), the authors of this paper acknowledge that the paper was reviewed, edited, and refined with the assistance of DeepL and Claude (Versions as of January 2026), complementing the human editorial process. The human authors critically assessed and validated the content to maintain academic rigor. The authors also assessed and addressed potential biases inherent in the AI-generated content. The final version of the paper is the sole responsibility of human authors.

References

1. London, A.J., Chang, M.L., Reig, S. *et al.* Avoiding a bridge to nowhere: managing the transfer of agency when an older adult can no longer use assistive AI. *AI Ethics* 6, 138 (2026). https://doi.org/10.1007/s43681-025-00981-5
2. Schroeder, T., Dodds, L., Georgiou, A., Gewald, H., & Siette, J. (2023). Older adults and new technology: Mapping review of the factors associated with older adults' intention to adopt digital technologies. *JMIR aging*, *6*(1), e44564.https://doi.org/10.2196/44564
3. Chen, G., Bhattacharya, S., & Spaulding, R. (2025). The role of AI driven personal assistants in geriatric care: opportunities, challenges, and future directions. *Ann Gerontol Geriatr Res*, *1*(1), 1003. https://annggr.org/articles/GGR-1003.html
4. Cao, X., Zhang, H., Zhou, B., Wang, D., Cui, C., & Bai, X. (2024). Factors influencing older adults' acceptance of voice assistants. *Frontiers in Psychology*, *15*, 1376207. https://doi.org/10.3389/fpsyg.2024.1376207
5. Jones, V. K., Yan, C., Shade, M. Y., Boron, J. B., Yan, Z., Heselton, H. J., ... & Dube, V. (2024). Reducing loneliness and improving social support among older adults through different modalities of personal voice assistants. *Geriatrics*, *9*(2), 22. doi: 10.3389/fpubh.2021.750736
6. Quinn, K., Leiser Ransom, S., O'Connell, C., Muramatsu, N., Marquez, D. X., & Chin, J. (2024). Assessing the feasibility and acceptability of smart speakers in behavioral intervention research with older adults: mixed methods study. *Journal of medical Internet research*, *26*, e54800. https://doi.org/10.2196/54800
7. Benevento, A. D., Ciulla, G., & Merlo, G. (2025). Future technologies in alternative and augmented communication: a scoping review of innovations. *Frontiers in Communication*, *10*, 1607531. https://doi.org/10.3389/fcomm.2025.1607531
8. Kim S. Exploring how older adults use a smart speaker-based voice assistant in their first interactions: qualitative study. JMIR Mhealth Uhealth. 2021;9(1): e20427. https://doi.org/10.2196/20427
9. Portet, F., Vacher, M., Golanski, C. *et al.* Design and evaluation of a smart home voice interface for the elderly: acceptability and objection aspects. *Pers Ubiquit Comput* **17**, 127–144 (2013). https://doi.org/10.1007/s00779-011-0470-5
10. Reis, A., Paulino, D., Paredes, H., Barroso, J. (2017). Using Intelligent Personal Assistants to Strengthen the Elderlies' Social Bonds. In: Antona, M., Stephanidis, C. (eds) Universal Access in Human–Computer Interaction. Human and Technological Environments. UAHCI 2017. Lecture Notes in Computer Science(), vol 10279. Springer, Cham. https://doi.org/10.1007/978-3-319-58700-4_48
11. Pradhan, A., Mehta, K., & Findlater, L. (2018, April). " Accessibility Came by Accident" Use of Voice-Controlled Intelligent Personal Assistants by People with Disabilities. In *Proceedings of the 2018 CHI Conference on human factors in computing systems* (pp. 1-13). https://doi.org/10.1145/3173574.3174033
12. Şener, B. S., & Özkaya, H. (2023). Investigation of the Relationship Between Social Media Addiction and Orthorexia Nervosa in Adult Individuals Who Applied to Obesity Polyclinic. *Addicta: The Turkish Journal on Addictions*, *10*(2), 135-141. https://doi.org/10.5152/ADDICTA.2023.23042
13. J. Lee, S. Lee, C. -H. Kim and J. Yoon, "Technology-Transferability Analysis of Universities and Public Research Institutes Using Deep Neural Networks," in *IEEE Access*, vol. 11, pp. 135196-135211, 2023, doi: 10.1109/ACCESS.2023.3337830.
14. Katz, E., Blumler, J. G., & Gurevitch, M. (1973). Uses and Gratifications Research. *The Public Opinion Quarterly*, *37*(4), 509–523. http://www.jstor.org/stable/2747854
15. Rauschnabel, P. A. (2018). Virtually enhancing the real world with holograms: An exploration of expected gratifications of using augmented reality smart glasses. *Psychology & Marketing*, *35*(8), 557-572. https://doi.org/10.1002/mar.21106

16. Kumar, P., Chauhan, S., & Awasthi, L. K. (2023). Artificial intelligence in healthcare: review, ethics, trust challenges & future research directions. *Engineering Applications of Artificial Intelligence*, *120*, 105894. https://doi.org/10.1016/j.engappai.2023.105894
17. Athwal N, Istanbulluoglu D, McCormack SE (2019), "The allure of luxury brands' social media activities: a uses and gratifications perspective". *Information Technology & People*, Vol. 32 No. 3 pp. 603–626. https://doi.org/10.1108/ITP-01-2018-0017
18. Elmas, E. (2024). *Kullanımlar ve Doyumlar Teorisi Perspektifinden Instagram Fenomenlerinin Marka Etkileşimi Üzerindeki Rolü ve bir Araştırma* (Unpublished master's thesis). Marmara University.
19. Creswell JW, Clark VLP. Designing and conducting mixed methods research. Sage Publications; 2017.
20. Patton C, Sawicki D, Clark J. Basic methods of policy analysis and planning. Routledge; 2015.
21. Palinkas, L.A., Horwitz, S.M., Green, C.A. *et al.* Purposeful Sampling for Qualitative Data Collection and Analysis in Mixed Method Implementation Research. *Adm Policy Ment Health* 42, 533–544 (2015). https://doi.org/10.1007/s10488-013-0528-y
22. Sayago, S. (Ed.). (2019). *Perspectives on human-computer interaction research with older people*. Switzerland: Springer International Publishing.
23. Seaborn, K., Sawa, Y., & Watanabe, M. (2024). Coimagining the future of voice assistants with cultural sensitivity. *Human Behavior and Emerging Technologies*, *2024*(1), 3238737. https://doi.org/10.1155/2024/3238737
24. Zhu, D., Al Mahmud, A., & Liu, W. (2024). Digital storytelling intervention for enhancing the social participation of people with mild cognitive impairment: co-design and usability study. *JMIR aging*, *7*, e54138.https://doi.org/10.2196/54138
25. Kvale S, Brinkmann S. InterViews: Learning the craft of qualitative research interviewing. SAGE Publications; 2015.
26. Rudhumbu, N. Antecedents of university lecturers' intentions to adopt information and communication technology in Zimbabwe. *Educ Inf Technol* **25**, 5117–5132 (2020). https://doi.org/10.1007/s10639-020-10205-4
27. Nielsen J. Usability 101: Introduction to usability [Internet]. Nielsen Norman Group; 2012. Available from: https://www.nngroup.com/articles/usability-101-introduction-to-usability/
28. Angrosino, M. (2007). *Doing ethnographic and observational research*. Sage.
29. Creswell, J. W., & Poth, C. N. (2018). Qualitative Inquiry and Research Design: Choosing among five approaches. SAGES Publication.
30. Charness N, Boot WR. Aging and Information Technology Use: Potential and Barriers. Curr Dir Psychol Sci. 2009;18(5):253–8.https://doi.org/10.1111/j.1467-8721.2009.01647.x
31. Kowalski J, Jaskulska A, Skorupska K, Abramczuk K, Biele C, Kopeć W, et al. Older adults and voice interaction: A pilot study with Google Home. In: Proceedings of CHI EA 2019. ACM; 2019. p. 1–6.https://doi.org/10.1145/3290607.3312973
32. Sedikides C., Wildschut T., Routledge C., and Arndt J. (2015) Nostalgia counteracts self-discontinuity and restores self-continuity, Eur. J. Soc. Psychol., 45, 52–61, doi: 10.1002/ejsp.2073.
33. Mihailidis, A., Boger, J.N., Craig, T. *et al.* The COACH prompting system to assist older adults with dementia through handwashing: An efficacy study. *BMC Geriatr* **8**, 28 (2008). https://doi.org/10.1186/1471-2318-8-28
34. Galliers, J. R. & Wilson, S. (2013). *An Exploratory Study into the Accessibility of a Multi-User Virtual World for Young People with Aphasia*. Paper presented at the BCS HCI 2013 - The Internet of Things XXVII, 9 - 13 Sep 2013, Brunel University, Uxbridge, UK. http://ewic.bcs.org/content/ConWebDoc/51712
35. Portet, F., Vacher, M., Golanski, C. *et al.* Design and evaluation of a smart home voice interface for the elderly: acceptability and objection aspects. *Pers Ubiquit Comput* **17**, 127–144 (2013). https://doi.org/10.1007/s00779-011-0470-5

36. Seaborn, K., Sawa, Y., & Watanabe, M. (2024). Coimagining the future of voice assistants with cultural sensitivity. Human Behavior and Emerging Technologies, 2024, 1–21. https://doi.org/10.1155/2024/1
37. Kowalski, J., Jaskulska, A., Skorupska, K., Abramczuk, K., Biele, C., & Kopeć, W. (2019). Older adults and voice interaction: A pilot study with Google Home. In Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (pp. 1–6). ACM.https://doi.org/10.1145/3290607.3312973
38. Kebede, A. S., Ozolins, L. L., Holst, H., & Galvin, K. (2022). Digital engagement of older adults: scoping review. *Journal of Medical Internet Research*, *24*(12), e40192. https://doi.org/10.2196/40192
39. Miles, M. B., & Huberman, A. M. (1994). Qualitative data analysis: An expanded sourcebook (2nd ed.). Sage Publications.
40. Prinos, K., Patwari, N., & Power, C. A. (2024). Speaking of accent: A content analysis of accent misconceptions in ASR research. In Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (pp. 1245–1254). Association for Computing Machinery. https://doi.org/10.1145/3630106.3658969
41. Kim, S., & Choudhury, A. (2021). Exploring older adults' perception and use of smart speaker-based voice assistants: A longitudinal study. Computers in Human Behavior, 124, 106914. https://doi.org/10.1016/j.chb.2021.106914
42. Liu, M., Wang, C., & Hu, J. (2023). Older adults' intention to use voice assistants: Usability and emotional needs. Heliyon, 9(11), e21932. https://doi.org/10.1016/j.heliyon.2023.e21932
43. Zhong, R., & Ma, M. (2022). Effects of communication style, anthropomorphic setting and individual differences on older adults using voice assistants in a health context. BMC Geriatrics, 22, 751. https://doi.org/10.1186/s12877-022-03428-2
44. Bozkurt, A. (2024). GenAI et al.: Cocreation, Authorship, Ownership, Academic Ethics and Integrity in a Time of Generative AI. Open Praxis, 16(1), pp. 1–10. https://doi.org/10.55982/openpraxis.16.1.654
45. Liu, M., Wang, C., & Hu, J. (2023). Older adults' intention to use voice assistants: Usability and emotional needs. *PubMed Central*, *9*(11). https://doi.org/10.1016/j.heliyon.2023.e21932

UNEC Journal of Computer Science and Digital Technologies
Volume 2, № 1, 2026, pp. 56-65
https://doi.org/10.30546/UNECCSDT.2026.001.288

Section:
Human–Computer Interaction &
Digital Education Technologies

AI-Based Personalized Learning and Student Performance: Evidence from Indian Higher Education

Muruganandham S K
Computer Science and Design Engineering, Raja Rajeswari College of Engineering, Bengaluru, Karnataka, India
0009-0000-7412-4178, drmurugax@gmail.com

Abstract

While the number of AI-powered educational platforms being developed and widely adopted in higher education institutions in India is rapidly increasing, there is not a lot of empirical evidence to determine whether such systems are effective for enhancing student outcomes. The study involved giving a structured quantitative questionnaire of the selected colleges of India to undergraduate and postgraduate students. Out of the 506 responses obtained, there were 497 responses that were retained after informed consent screening. The instrument measured four constructs related to AI-based personalized learning, students' learning engagement, and their satisfaction and performance using a 24-item, five-point Likert-scale. Descriptive statistics, Pearson correlation, multiple regression, and bootstrapped mediation analysis were conducted. Personalized learning based on AI, along with engagement and satisfaction, accounted for 75.1 % of the variance in academic performance (R2 = 0.751, p < 0.001), with both mediators shown to be significant. The results provide empirical insights on how to guide institutions, policymakers, and EdTech companies to reinforce learning outcomes through AI for personalisation in Indian higher education.

Keywords: Artificial Intelligence, Personalized Learning, Academic Performance, Learning Engagement, Student Satisfaction, Indian Higher Education

Received:	*Revised:*	*Accepted:*	*Published:*
29/05/2026	*03/06/2026*	*06/06/2026*	*17/06/2026*

1. Introduction

While much investment has been made in higher education over the last decades, the one size fits all classroom approach to teaching continues to be the reality. This is starting to change with the use of AI in learning platforms: AI systems are now capable of identifying how well a student is doing at each moment and adapting the content to that level, and of suggesting learning routes to take based on what the student has already shown their ability. What the study will tell us is whether or not this ability will actually lead to a boost in student achievement. It is all the more pressing in the Indian context. As per the University of the States and University Grants Commission (UGC), there are 1006 universities and 42,022 colleges in India enrolling more than 43 million students in various courses, resulting in one of the largest and most diverse higher education systems in the world, which is still facing challenges of shortage of qualified faculty, infrastructure gaps, and varying levels of student preparation [21]. Using platforms like BYJU'S, Unacademy, upGrad or Vedantu, hundreds of millions of students are able to learn, and the number of Indians accessing these services has increased significantly during the pandemic of COVID-19 [10]; however, little outcome research has matched the rapid growth in uptake. Previous studies of the adoption in India are about the intention to adopt or perceptions of achievement [22, 25] and few (if any) investigated learning engagement and satisfaction as mediators between personalization–performance.

This study contributes to filling this gap. This survey was designed to use a 24 item, 5-point Likert scale that was given to the undergraduate and postgraduate students of selected Indian Institutions (N = 506, n = 497). Seven hypotheses based on Technology Acceptance Model [7], Self-Determination Theory [8] and IS Success Model [9] were tested using multiple regression and bootstrapped mediation analysis. The four goals for the work were: (1) capture the usage patterns with the AI platform; (2) explore the potential direct impact of AI personalization on performance; (3) identify the connection between the AI platform and performance through the mediating factors of engagement and satisfaction; and (4) draw implications for the institution, curriculum designers, and EdTech providers. The paper continues by conducting literature review, presenting the theoretical framework, presenting the methodology, presenting the results, conducting discussion and then a conclusion.

2. Literature Review

Bloom's two-sigma discovery [1] - that AI tutoring at scale is capable of improving learning by 2 sigmas compared to conventional learning - set the ceiling that AI based personalized learning aspires to. The results of the meta-analysis strongly support the efficacy of intelligent tutoring systems, with average achievement gains of 0.4–0.8 compared to traditional teaching and learning [19,27,29] but this can vary depending on the quality of the implementation and the characteristics of the student population [5]. In contrast to the first generation of tutoring systems, state-of-the-art AI approaches automatically adapt to individual learners and adapt the sequencing of content, the timing of feedback, and the difficulty level of the assessments on the fly, and are based on continually updated representations of each student [24, 26]. Learning engagement is one of the most powerful ones to predict learning outcomes [14] and is the mechanism through which AI personalization is theorised to yield better performance learning across behavioural, cognitive and emotional levels [12]. Adaptive platforms can keep learning opportunities challenging within each learner's zone of proximal development (ZPD) [30] and maintain effortful processing opportunities for learning gains. Huang2020 (in a sample of a university) confirmed much higher engagement with the use of adaptive systems; Filgona2020 found that engagement was a full mediator on this technology-achievement link and Panigrahi2018 found similar in Indian e-learning contexts. Student satisfaction is an alternative course of action. Satisfaction has been found to be a motivational mechanism to achieve better performance outcomes as found in the Technology Acceptance Model [7] or IS Success Model [9] in Taiwan [28], in Brazil [3] and in the United States [18].

Table 1. *The prior work*

Author(s)	Year	Country	Method	Key Finding
Bloom [1]	1984	USA	Experiment	Tutoring exceeds conventional instruction by ≈2SD
VanLehn [29]	2011	USA	Meta-anal.	AI tutoring approaches human effectiveness; d≈0.76
Ma et al. [19]	2014	Global	Meta-anal.	Significant ITS gains across subjects and grade levels
Sun et al. [28]	2008	Taiwan	Survey	Satisfaction predicts e-learning outcomes
Cidral et al. [3]	2018	Brazil	SEM	Satisfaction mediates quality to performance link
Panigrahi et al. [25]	2018	India	Survey	TAM predicts e-learning adoption in Indian HEIs

Dhawan [10]	2020	India	Qualitative	EdTech potential post-COVID; outcome gap noted
Huang et al. [16]	2020	China	Survey	Adaptive systems raise student engagement
Filgona et al. [11]	2020	Nigeria	Quasi-exp.	Engagement mediates technology–achievement link
modi2025 & insights [22]	2021	India	Survey	Positive AI tool perceptions in Indian engineering colleges

Whilst India's National Education Policy 2020 has set AI personalization as a key direction to pursue in the field of education [20] and a small domestic EdTech industry is expected to be over a USD 10 billion market by 2025 [17], there is poor interest in rigorous outcome research in higher education in India. There have been existing studies that have looked into adoption intentions [2] or perceptions but with no measured performance [10,22]. This is covering the missing area in the existing literature as none of these studies simultaneously look at engagement/satisfaction as mediators in the personalization–performance relationship in an Indian undergraduate and postgraduate sample. The prior work is summarized in key aspects in Table 1. The concept model is supported by 3 complementary frameworks. The Technology Acceptance Model (TAM)[7] is a theory that states that perceived usefulness and perceived ease of use are the main factors to encourage satisfaction and continuous engagement on the platform. The "why?" behind AI personalization that deepens learning is explained in Self-Determination Theory [8]: Sustaining challenge with each learner's zone of proximal development fosters the three basic needs for intrinsic motivation and deeper cognitive processing, namely autonomy, competence and relatedness. According to IS Success Model [9] there is a relationship between system quality and satisfaction which can be linked to net benefits which in this study is academic performance through satisfaction. The three frameworks together suggest that personalisation with AI has a 180-degree effect on performance, in direct and indirect (mediated) ways, both with regard to engagement and satisfaction. Figure 1 shows the conceptual model that was integrated into the developed conceptual model, and Table 2 shows that 7 hypotheses were stated in accordance with the integrated conceptual framework.

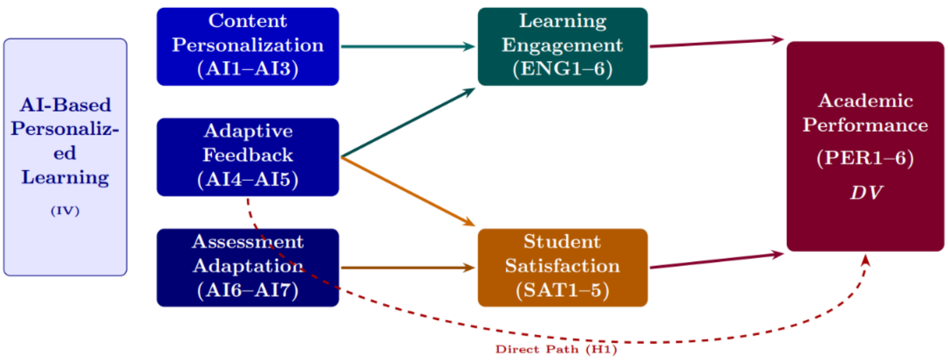

Figure 1. Integrated conceptual framework. Solid arrows = hypothesized paths; dashed = direct path (H1). TAM [7]; SDT [8]; IS Success Model [9]

3. Research Methodology

3.1 Research Design and Sampling

The study is a quantitative, cross section survey [6]. For undergraduate and post-graduate students in India Higher Education Institutes (HEI) with direct experience of using AI learning platform were targeted. A combination of purposive samples with the snowball sample (using academic networks, departmental messaging groups, and institutional e-mail lists was used to improve the chances of respondents having the needed experience). The survey was conducted online using google forms. From a total of 506 returned questionnaires, 13 were omitted who refused to provide informed consent, and there were 497 valid questionnaires which is slightly more than the minimum amount of 384 questionnaires obtained by Cochran's formula [4] with regard to informed consent (confidence level: 95%, margin of error: 5%) which is the power needed for regression and bootstrapped mediation analysis [13]. The procedure for screening is shown in the figure 2

Table 2. *Research Hypotheses Summary*

Hyp.	Statement	Path Type	Dir.
H1	AI personalization directly influences academic performance	Direct	+
H2	AI personalization positively affects learning engagement	Direct	+
H3	Learning engagement positively affects academic performance	Direct	+
H4	AI personalization positively affects student satisfaction	Direct	+
H5	Student satisfaction positively affects academic performance	Direct	+
H6	Engagement mediates AI personalization and performance	Mediation	+
H7	Satisfaction mediates AI personalization and performance	Mediation	+

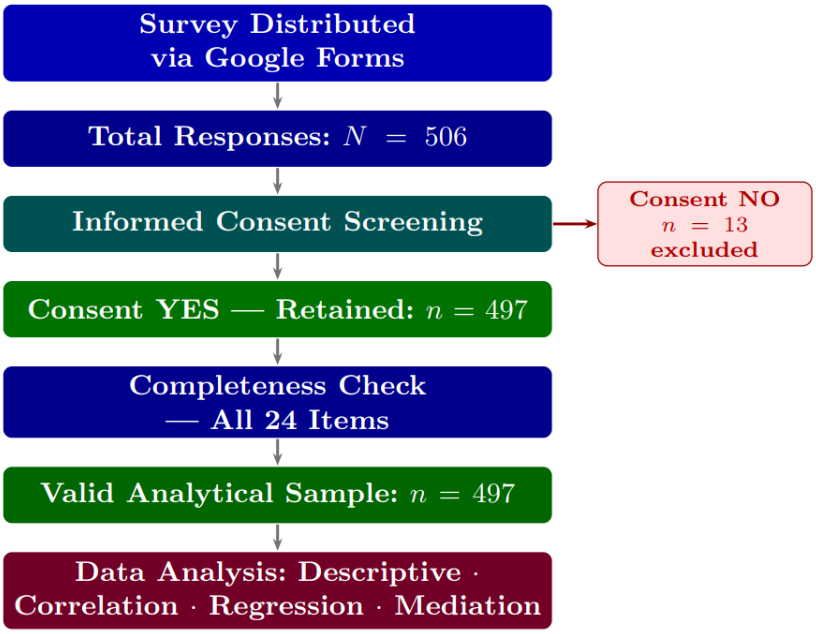

Figure 2. Data collection and screening procedures.

3.2 Measurement Instrument

The instrument consisted of 24 items which were given a 5-point Likert scale (1 = Strongly Disagree, 2 = Disagree, 3 = Don't Know, 4 = Agree, 5 In each of the 4 constructs the (SAH) strongly agree indicated that the item received high agreement responses. Personalized learning items with the use of AI. The technology affection (AI1 to AI7) items were based on the items from Davis [7] and Pane et al. [24], the learning engagement items (ENG1 for ENG6) were based on the items from Fredricks et al. [12] and the student satisfaction items (SAT1 to SAT5) were based on the items from DeLone and McLean [9] and the academic performance items (PER1 to PER6) were based on the items from Cidral et al. [3] and Kuo et al. [18]. All items were reviewed by three ET faculty members to ensure their clarity and relevance and were pilot tested with 30 students who found them to be acceptable and showed good levels of reliability (α > 0.80) with some minor changes to items. A final Cronbach's alpha ranged from 0.898 to 0.929 and indicated a good coefficient for each of the four scales (> 0.70) [23]. All of the 24 items are listed in table 3.

Table 3: *Measurement Instrument — All 24 Scale Items*

Code	Item
AI-Based Personalized Learning (IV, 7 items) — Davis [7]; Pane et al. [24]	
AI1	The AI platform adapts learning content to my individual needs
AI2	The system recommends materials based on my learning performance
AI3	The learning pace is adjusted according to my understanding level
AI4	The AI system provides customized feedback on my performance
AI5	The content provided by the AI system matches my learning preferences
AI6	The system identifies my weak areas and suggests improvement strategies
AI7	The AI platform personalizes assessments based on my progress
Learning Engagement (Mediator, 6 items) — Fredricks et al. [12]	
ENG1	I feel actively involved when using AI-based learning systems
ENG2	AI-based learning increases my interest in the subject
ENG3	I spend more time learning when using AI-personalized tools
ENG4	The AI system motivates me to participate actively in learning tasks
ENG5	I concentrate better when learning through AI-personalized systems
ENG6	AI-based platforms make learning more interactive
Student Satisfaction (Mediator, 5 items) — DeLone & McLean [9]	
SAT1	I am satisfied with the AI-based personalized learning experience
SAT2	The AI system meets my learning expectations
SAT3	I find AI-based learning more effective than traditional methods
SAT4	I would recommend AI-based personalized learning to others
SAT5	I am pleased with the overall performance of AI learning platforms
Academic Performance (DV, 6 items) — Cidral et al. [3]; Kuo et al. [18]	
PER1	My academic performance has improved due to AI-based personalized learning
PER2	I understand concepts better when using AI-personalized platforms
PER3	My grades have improved after using AI learning systems
PER4	AI-based learning has increased my learning efficiency
PER5	I complete assignments more effectively using AI tools
PER6	AI-based personalized learning has enhanced my overall academic achievement

3.3 Data Analysis

Analysis proceeded in five stages: (1) descriptive statistics for all items and demographics; (2) Cronbach's alpha per scale; (3) Pearson correlations among construct scores with multicollinearity check ($r < 0.90$ criterion); (4) simultaneous multiple regression of academic performance on all three predictors; and (5) bootstrapped mediation analysis (B = 5,000 resamples, 95 % bias-corrected CI) testing H6 and H7 [15]. All analyses were conducted in Python using pandas, NumPy, and SciPy.

4 Results

4.1 Sample Profile

The sample (n = 497) was 52.1 % female, 55.5 % aged 18–20, and 75.5 % undergraduate. Engineering and Technology students dominated (87.3 %), with 97.6 % from private institutions. Over 83 % engaged with AI learning tools at least sometimes, with ChatGPT and similar chatbots in the most widely used category. Table 4 presents the full breakdown.

4.2 Descriptive Statistics and Reliability

Construct means ranged from 3.764 (student satisfaction) to 3.893 (AI personalization) on the five-point scale, indicating broadly positive perceptions. Cronbach's alpha ranged from 0.898 to 0.929, all exceeding the 0.70 threshold [23] and classified as excellent. Table 5 reports full statistics.

Table 4. *Demographic Profile (n = 497)*

Variable	Category	n	%
Gender	Female	259	52.1
	Male	232	46.7
	Prefer not to say	6	1.2
Age Group	18–20 years	276	55.5
	21–23 years	216	43.5
	24–26 years	5	1.0
Level of Study	Undergraduate	375	75.5
	Postgraduate	122	24.5
Institution Type	Private	485	97.6
	Government	12	2.4
AI Usage Frequency	Sometimes	158	31.8
	Very Frequently	130	26.2
	Often	127	25.6
	Occasionally	46	9.3
	Rarely	36	7.2

Table 5. *Descriptive Statistics and Reliability (n = 497)*

Construct	Items	M	SD	Min	Max	α
AI Personalization (IV)	7	3.893	0.785	1.00	5.00	0.898
Learning Engagement	6	3.809	0.864	1.00	5.00	0.921
Student Satisfaction	5	3.764	0.859	1.00	5.00	0.901
Academic Performance (DV)	6	3.798	0.889	1.00	5.00	0.929

Note. Scale: 1 = Strongly Disagree, 5 = Strongly Agree

4.3 Correlation Analysis

All pairwise correlations were positive and significant ($p < 0.01$), ranging from $r = 0.681$ (AI personalization – performance) to $r = 0.837$ (satisfaction – performance), all below the 0.90 multicollinearity threshold. The comparatively lower direct correlation for AI personalization signals the importance of the mediating pathways. Table 6 reports the matrix.

Table 6. *Pearson Correlation Matrix (n = 497)*

Construct	1	2	3	4
1. AI Personalization	1.000			
2. Learning Engagement	0.735**	1.000		
3. Student Satisfaction	0.701**	0.816**	1.000	
4. Academic Performance	0.681**	0.811**	0.837**	1.000

** $p < 0.01$, two-tailed

4.4 Multiple Regression Analysis

The model was significant ($F(3, 493) = 496.93$, $p < 0.001$) and explained 75.1 % of variance in academic performance ($R2 = 0.751$, Adj. $R2 = 0.750$). Student satisfaction was the strongest predictor (β = 0.503, $p < 0.001$), followed by learning engagement (β = 0.347, $p < 0.001$). The direct effect of AI personalization remained significant after controlling for both mediators (β = 0.073, p = 0.034), confirming H1. Full results are in Table 7.

Table 7. *Multiple Regression Results — DV: Academic Performance (n = 497)*

Predictor	B	SE	β	t	p
Constant	0.159	0.104	—	1.525	0.128
AI Personalization	0.083	0.039	0.073	2.130	0.034*
Learning Engagement	0.357	0.044	0.347	8.207	<0.001**
Student Satisfaction	0.520	0.042	0.503	12.516	<0.001**

$*p < 0.05$; $**p < 0.001$

4.5 Mediation and Path Analysis

Both indirect effects were significant. The effect of AI personalization on academic performance via learning engagement was IE = 0.561 (95 % BC CI [0.471, 0.653]), and via student satisfaction was IE = 0.561 (95 % BC CI [0.476, 0.648]). Both intervals exclude zero, confirming H6 and H7 [15]. All seven hypotheses were supported. Table 8 and Figure 3 summarise the full path structure.

Table 8. *Bootstrapped Mediation Analysis (B = 5,000, n = 497)*

Mediator (Hyp.)	Indirect	Effect	95 % BC CI
Learning Engagement (H6)	0.561	Significant	[0.471, 0.653]
Student Satisfaction (H7)	0.561	Significant	[0.476, 0.648]

Note. BC CI excludes zero ⇒ mediation confirmed [15].

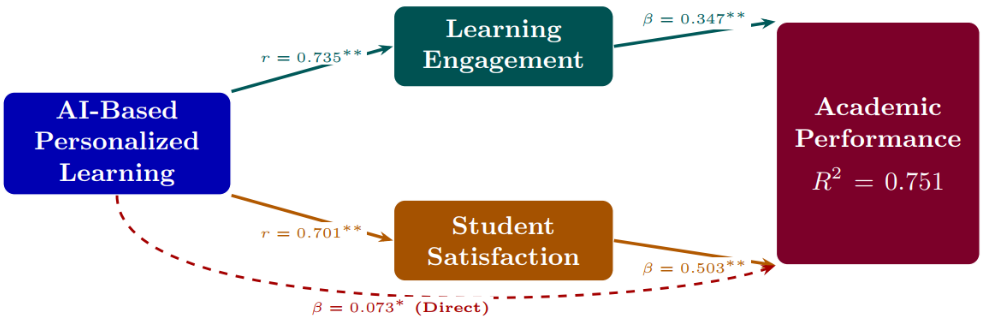

Figure 3. Structural path diagram with standardised coefficients. $*p < 0.05$; $**p < 0.001$. Dashed = direct path after controlling for mediators.

5. Discussion

This three-predictor model accounted for a high (for such a parsimonious model) 75.1 % of the variance in academic performance and all seven hypotheses were supported. This large direct effect of AI Personalization (β = 0.073, p = 0.034) after taking into account both mediators, aligns with the findings of VanLehn [29] and Ma et al. [19] that AI Personalization results in a direct effect on structural performance which is not mediated through motivational effects. The finding of this effect in an Indian sample where quality of platforms, access to devices and digital literacy are quite different, portrays the strength of the phenomenon at the underlying level. But it is the mediated pathways' sizes that are the prevailing finding. The indirect effects (IE = 0.561) are much higher, making it clear that AI personalization helps to boost performance by increasing engagement levels and satisfaction as well as by directly delivering content. The engagement pathway is consistent with Fredricks et al [12] who relate this to Fredricks et al [13] 'productive challenge within each person's ZPD [30] maintaining the behavioural, cognitive, and emotional engagement relates to greater learning gains. This is similar to the discovery in China [16] and in Nigeria [11] where the engagement mediation structure was discovered in the educational space of the HEIs in South Asia. Consistent with the IS Success Model [9] – reproduced in Taiwan [28] and Brazil [3] – the results confirmed, as predicted by TAM [7] that perceived usefulness translates into consistent and effortful use which in turn can lead to an increase in learning gains over time. The higher β for satisfaction of 0.503 than that in similar past studies might be due to the unique importance of affective evaluations given the relative novelty of using AI tools, particularly for students. There is a high correlation between engagement and satisfaction ($r = 0.816$), but they do regress independently and significantly on performance, so they do seem to be reflecting genuinely different variances in performance. Interventions in either path alone (just engagement or just satisfaction) will be limited in the increments to performance resulting. The pathway of integrating TAM–SDT–IS Success framework [7-9] could explain 75.1 % of outcome variance, which indicated that none of these theories is sufficient for explaining the outcome and the combination of these theories is suitable in AI-based learning study. Caution should be used in generalisation when using the composition of samples. However, students under control of engineering/technology (87.3 %) at private educational institutes (97.6 %) are likely to be more digitally fluent and better resourced, as compared to the overall student population in the country [21]. The findings are pertinent to this sizeable and expanding user group – one of the highest levels of EdTech use in the country – however, if NEP 2020's hope of broad rolling out of AI personalization throughout India's entire HE system is to be realised, evidence from contexts such as arts and humanities, government funded and rural institution contexts must also be considered.

6. Conclusion

Survey data from 497 Indian undergraduate and postgraduate students indicate that when we offer personalized learning based on AI, it not only directly enhances academic performance ($\beta = 0.073$, $p = 0.034$), but also indirectly via the mediators learning engagement (IE = 0.560, 95 % CI [0.470, 0.652]) and student satisfaction (IE = 0.560, 95 % CI [0.475, 0.647]). The three-predictor model explained 75.1 % of outcome variance (R2 = 0.751, $F(3, 493) = 496.93$, $p < 0.001$), with all seven hypotheses supported. The main performance improvement comes as a result of this increased engagement and satisfaction and not simply the provision of content on its own - which reinforces the mediation structures identified in other studies in China [16] and Nigeria [11] and demonstrates their generalizability to the South Asian context. This structure is represented in the integrated TAM – SDT-IS success framework [7-9] which has two separate, although related and non-redundant, explanatory pathways coming from TAM and SDT.This has implications that are far and out there and not just sidewalks. AI learning programmes need to shift from the fringes to the core, with a focus on the pedagogical curriculum of faculty orientation, student orientation and usability of the platform. Ban on the use of artificial intelligence or not, service designers and developers in the EdTech space should focus on satisfaction as a key performance metric – because residences of pupils on this metric on SAT3 (M = 3.680) suggests that there is still some degree of uncertainty that AI doesn't have an advantage over traditional approaches, which can be resolved through good design and communication. Policy makers pursuing NEP 2020's vision of AI personalisation should also focus on providing infrastructure and connectivity, which are prerequisites to engagement and satisfaction as mechanisms to achieving the benefits. These conclusions build on four limitations: the cross-sectional design does not allow for causal inferences; all constructs are self-reported; the sample is heavily skewed towards Engineering students from private schools and their ability to generalize to other fields and across all schools is limited; and, AI personalization is considered one construct, even though practice varies across platforms. Future research will include longitudinal or experimental research, increasing beyond populations in Arts/Humanities and Government supported institutions, and disaggregating results based on platform type, while including rural institution populations. Then sixty years after the insightful two sigma gap identified by Bloom [1] the question is, whether AI personalization will be able to do so in India or not will primarily depend on how much these three stakeholders viz., institutions, developers, and policymakers will take cognizance of the human factors that this study highlights.

Acknowledgements

The author thanks the students who participated in this survey and the faculty colleagues who reviewed the instrument prior to deployment

Authors' Declaration Conflicts of Interest: The author declares no conflict of interest.

Authors' Contribution Statement:

Muruganandham SK: Conceptualization, Methodology, Data collection, Formal analysis, Investigation, Writing - original draft, Writing - review & editing

References

1. Bloom, B.S.: The 2-sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. Educational Researcher 13(6), 4–16 (1984). https://doi.org/10.3102/0013189X013006004
2. Chen, J., Chang, T.: Exploring factors influencing AI-powered e-learning system adoption intention: An empirical study on mediation and moderation effects. Environment and Social Psychology 10(3) (2025). https://doi.org/10.59429/esp.v10i3.3551
3. Cidral, W.A., Oliveira, T., Di Felice, M., Aparicio, M.: E-learning success determinants: Brazilian empirical study. Computers & Education 122, 273–290 (2018). https://doi.org/10.1016/j.compedu.2017.12.001
4. Cochran, W.G.: Sampling Techniques. John Wiley & Sons, New York, USA, 3rd edn. (1977)
5. Cohen, J.: Statistical Power Analysis for the Behavioral Sciences. Lawrence Erlbaum Associates, Hillsdale, NJ, 2nd edn. (1988)
6. Creswell, J.W.: Research Design: Qualitative, Quantitative, and Mixed Methods Approaches. SAGE Publications, Thousand Oaks, CA, 4th edn. (2014)
7. Davis, F.D.: Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly 13(3), 319–340 (1989). https://doi.org/10.2307/249008
8. Deci, E.L., Ryan, R.M.: Intrinsic Motivation and Self-Determination in Human Behavior. Plenum Press, New York, USA (1985). https://doi.org/10.1007/978-1-4899-2271-7
9. DeLone, W.H., McLean, E.R.: The DeLone and McLean model of information systems success: A ten-year update. Journal of Management Information Systems 19(4), 9–30 (2003). https://doi.org/10.1080/07421222.2003.11045748
10. Dhawan, S.: Online learning: A panacea in the time of COVID-19 crisis. Journal of Educational Technology Systems 49(1), 5–22 (2020). https://doi.org/10.1177/0047239520934018
11. Filgona, J., Sakiyo, J., Gwany, D.M., Okoronkwo, A.U.: Motivation in learning. Asian Journal of Education and Social Studies 10(4), 16–37 (2020). https://doi.org/10.9734/ajess/2020/v10i430273
12. Fredricks, J.A., Blumenfeld, P.C., Paris, A.H.: School engagement: Potential of the concept, state of the evidence. Review of Educational Research 74(1), 59–109 (2004).

https://doi.org/10.3102/00346543074001059
13. Hair, J.F., Risher, J.J., Sarstedt, M., Ringle, C.M.: When to Use and How to Report the Results of PLS-SEM, vol. 31 (2019). https://doi.org/10.1108/EBR-11-2018-0203
14. Hattie, J.: Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement. Routledge, London, UK (2009)
15. Hayes, A.F.: Introduction to Mediation, Moderation, and Conditional Process Analysis. Guilford Press, New York, USA, 2nd edn. (2018)
16. Huang, R., Tlili, A., Chang, T.W., Zhang, X., Nascimbeni, F., Burgos, D.: Disrupted classes, undisrupted learning during the COVID-19 outbreak in China: Application of open educational practices and resources. Smart Learning Environments 7(1), 1–15 (2020). https://doi.org/10.1186/s40561-020-00125-8
17. KPMG India: Online education in India: 2021 market report. Tech. rep., KPMG and Google, Mumbai, India (2022), https://home.kpmg/in
18. Kuo, Y.C., Walker, A.E., Schroder, K.E.E., Belland, B.R.: Interaction, internet self-efficacy, and self-regulated learning as predictors of student satisfaction in online education. The Internet and Higher Education 20, 35–50 (2014). https://doi.org/10.1016/j.iheduc.2013.10.001
19. Ma, W., Adesope, O.O., Nesbit, J.C., Liu, Q.: Intelligent tutoring systems and learning outcomes: A meta-analysis. Journal of Educational Psychology 106(4), 901–918 (2014). https://doi.org/10.1037/a0037123
20. Ministry of Education, Government of India: National education policy 2020. Tech. rep., Ministry of Education, New Delhi, India (2020), https://www.education.gov.in/nep
21. Ministry of Education, Government of India: All india survey on higher education 2021–22. Tech. rep., Department of Higher Education, Ministry of Education, New Delhi, India (2023), https://aishe.gov.in
22. Modi, L., Salunkhe, K., Pant, M., Shree, S., Rote, S., Doke, R.: Insights from an online survey of higher education students across disciplines in maharashtra: Exploring the adoption and impact of AI-powered tools in learning. Journal of Information Systems Engineering and Management 10(4) (2025). https://doi.org/10.52783/jisem.v10i4.9830
23. Nunnally, J.C., Bernstein, I.H.: Psychometric Theory. McGraw-Hill, New York, USA, 3rd edn. (1994)
24. Pane, J.F., Steiner, E.D., Baird, M.D., Hamilton, L.S.: Continued Progress: Promising Evidence on Personalized Learning. RAND Corporation, Santa Monica, CA (2015). https://doi.org/10.7249/RR1365
25. Panigrahi, R., Srivastava, P.K., Sharma, D.: Online learning: Adoption, continuance, and learning outcome — a review of literature. International Journal of Information Management 43, 1–14 (2018). https://doi.org/10.1016/j.ijinfomgt.2018.05.005
26. Roll, I., Wylie, R.: Evolution and revolution in artificial intelligence in education. International Journal of Artificial Intelligence in Education 26(2), 582–599 (2016). https://doi.org/10.1007/s40593-016-0110-3
27. Steenbergen-Hu, S., Cooper, H.: A meta-analysis of the effectiveness of intelligent tutoring systems on K–12 students' mathematical learning. Journal of Educational Psychology 105(4), 970–987 (2013). https://doi.org/10.1037/a0032447
28. Sun, P.C., Tsai, R.J., Finger, G., Chen, Y.Y., Yeh, D.: What drives a successful E-learning? an empirical investigation of the critical factors influencing learner satisfaction. Computers & Education 50(4), 1183–1202 (2008). https://doi.org/10.1016/j.compedu.2006.11.007
29. VanLehn, K.: The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. Educational Psychologist 46(4), 197–221 (2011). https://doi.org/10.1080/00461520.2011.611369
30. Vygotsky, L.S.: Mind in Society: The Development of Higher Psychological Processes. Harvard University Press, Cambridge, MA (1978)

UNEC Journal of Computer Science and Digital Technologies
Volume 2, № 1, 2026, pp. 66-75
https://doi.org/10.30546/UNECCSDT.2026.001.256

ESG Indicators and Online Reputation in Tourism Platforms: A Pilot Data Analysis

Nijat Safarov
University of Economics in Katowice, Katowice, Poland
0009-0001-0550-6874, nijat.safarov@edu.uekat.pl

Abstract

This paper investigates whether sustainability performance, measured through environmental, social, and governance (ESG) indicators, is reflected in digital reputation signals on tourism platforms at the city level. Destination managers and public institutions increasingly rely on ESG-based metrics to evaluate sustainability. However, reputation systems on digital tourism platforms—such as ratings, reviews, and rankings—remain largely driven by perceived service quality and popularity. This disconnect may reduce transparency and limit the ability of platform-based demand to support sustainable destination governance. Using a pilot empirical design, the study links publicly available ESG indicators for fifteen tourism cities with platform-based reputation measures derived from attraction ratings and review volumes on TripAdvisor. The analysis examines relationships between ESG indicators and digital reputation signals, as well as differences across ESG dimensions and between evaluative indicators (average ratings) and visibility indicators (review volume). The results show that environmental and social indicators are modestly associated with perceived destination quality, while governance indicators are more strongly related to platform visibility. These findings provide preliminary empirical evidence on the ESG–reputation relationship and highlight the potential relevance of integrating sustainability indicators into digital tourism reputation systems.

Keywords: ESG indicators; tourism cities; online reputation; digital tourism platforms; sustainable destination governance

Received:	*Revised:*	*Accepted:*	*Published:*
20/05/2026	*01/06/2026*	*07/06/2026*	*17/06/2026*

1. Introduction

The emergence of digital tourism platforms has significantly transformed destination visibility and tourist decision-making processes. Reputation systems embedded in these platforms, such as ratings, reviews, and algorithmic rankings—function as key signals of quality and trust, influencing destination competitiveness and market outcomes in global tourism markets [1–3]. Although these mechanisms were originally designed to support engagement optimization and commercial performance, they increasingly operate as informal governance structures that shape how destinations are perceived and selected within digital tourism ecosystems [4].

At the same time, sustainability and responsible tourism development have become central priorities in destination management and public policy. Environmental, social, and governance (ESG) indicators are widely used by governments, international organizations, and institutional stakeholders to assess long-term sustainability performance and guide tourism development strategies [5–7]. These sustainability metrics aim to integrate tourism growth with environmental protection, social well-being, and effective governance practices. Recent research also emphasizes the growing importance of digital development processes and risk-oriented approaches in shaping contemporary digital ecosystems and governance environments [26, 27].

Despite this growing institutional emphasis on sustainability, ESG-based evaluations remain largely disconnected from the user-facing reputation systems embedded in digital tourism platforms. While platform reputation systems primarily reflect popularity and perceived service quality, they rarely capture broader environmental and social impacts associated with tourism activity [2, 8]. This disconnect creates a gap between institutional sustainability assessment and platform-mediated reputation formation.

To address this gap, the present study investigates whether sustainability performance, captured through ESG indicators, is associated with observable reputation signals on digital tourism platforms at the city level. Using a pilot data approach, the research links publicly available ESG indicators for selected tourism cities with platform-based reputation measures derived from user-generated evaluations. By empirically exploring the relationship between sustainability indicators and digital reputation signals, the study provides initial evidence on the potential integration of ESG-informed evaluation into platform reputation systems.

Recent empirical studies have examined the role of online reputation systems in shaping tourism demand and destination competitiveness, highlighting the influence of user-generated ratings and reviews on tourist decision-making [1–3, 8]. At the same time, a growing body of research has focused on sustainability and ESG-related indicators in tourism, emphasizing the importance of environmental quality, social conditions, and governance for long-term destination performance [5–7, 11–15]. However, empirical studies that directly link sustainability performance with platform-based reputation signals remain limited [13–17]. This study contributes to bridging this gap by providing an initial empirical examination of the relationship between ESG indicators and digital reputation at the city level.

Digital tourism platforms have become central infrastructures in contemporary tourism ecosystems, facilitating information exchange, shaping destination visibility, and influencing travel decisions worldwide [1, 8-10]. These platforms rely heavily on reputation systems based on user-generated ratings, reviews, and rankings to evaluate the trustworthiness, quality, and popularity of tourism destinations [2, 11–13]. However, the design of these reputation systems is primarily driven by visibility dynamics and market coordination, and they often face challenges related to transparency, credibility, and the potential manipulation of reviews [2, 9, 14–17].

At the same time, the increasing importance of Environmental, Social, and Governance (ESG) principles in tourism policy and destination management has transformed how tourism systems are evaluated. Sustainability frameworks increasingly rely on indicator-based approaches to measure environmental pressures, social well-being, governance quality, and long-term resilience of tourism destinations [18-23]. Despite the expansion of ESG-oriented evaluations, their connection to digital tourism platform reputation mechanisms remains poorly understood.

This lack of integration creates a structural governance gap. Reputation signals visible to travelers do not necessarily reflect the sustainability performance of destinations, while sustainability assessments used by institutions rarely translate into user-facing decision environments where tourism demand is formed in real time [18, 19, 24, 25]. Consequently, platform-mediated tourism demand may reinforce short-term popularity rather than long-term sustainable development.

Therefore, this study addresses the research question of whether measurable relationships exist between ESG-related sustainability indicators and online reputation signals in digital tourism platforms. By examining this relationship through a pilot empirical design at the city level, the study aims to provide an initial analytical basis for integrating sustainability performance into platform-based reputation systems and strengthening the governance role of digital tourism platforms in supporting sustainable destination development.

2. Experiments

2.1. Conceptual Model and Hypotheses

This study examines whether city-level sustainability performance, captured through Environmental, Social, and Governance (ESG) indicators, is reflected in observable online reputation signals on digital tourism platforms. Sustainability-related attributes such as environmental quality, safety conditions, and governance effectiveness may influence visitor experience and, consequently, user-generated evaluations of destinations.

However, reputation systems on digital tourism platforms are also shaped by visibility dynamics, popularity effects, and platform attention mechanisms. As a result, sustainability performance may not necessarily translate directly into a stronger digital reputation without explicit integration into platform evaluation systems.

Based on this reasoning, the study proposes the following hypotheses:

- H1 (overall association): Cities with higher overall ESG performance exhibit stronger online reputation on digital tourism platforms.
- H2 (dimension heterogeneity): The relationship between sustainability performance and online reputation differs across ESG dimensions (Environmental, Social, and Governance).
- H3 (signal-type heterogeneity): ESG performance relates differently to evaluative reputation signals (e.g., average attraction ratings) and visibility-based signals (e.g., review volume).

2.2 Data and Sample

2.2.1 Unit of analysis and sampling strategy

The unit of analysis in this study is the city. A pilot sample of 15 tourism cities was selected using a purposive-variation approach in order to capture differences in destination visibility, tourism maturity, and regional context. The final sample includes Paris, Rome, Dubai, Barcelona, Athens, Krakow, Vienna, Baku, Tbilisi, Yerevan, Samarkand, Astana, Istanbul, Cairo, and Marrakech.

The sample combines globally established destinations with emerging or regionally prominent tourism cities, thereby increasing variation in both online reputation indicators and sustainability-related conditions. This diversity enables an exploratory comparison of destinations with different levels of digital tourism visibility. The selection of cities follows a purposive sampling strategy aimed at maximizing variation across key dimensions relevant to the study. Specifically, the sample includes cities with different levels of tourism intensity (highly visited global destinations versus emerging or regionally significant cities), varying degrees of digital platform visibility (as reflected in review activity), and diverse geographic and institutional contexts. This approach allows the study to capture heterogeneity in both sustainability conditions and online reputation signals, which is necessary for exploring potential relationships between ESG performance and platform-based visibility and evaluation. The inclusion of both globally dominant and less internationally visible destinations enables comparison between cities with high platform visibility and those with lower digital prominence.

2.2.2 ESG indicators and data sources

City-level sustainability performance is measured using publicly available ESG-related data. It is operationalized using three comparable indicators: environmental quality (proxied by the Numbeo Pollution Index), social conditions (proxied by the Numbeo Safety Index), and governance effectiveness (derived from the World Bank Worldwide Governance Indicators).

For each city, the following indicators are recorded:

- Environmental (E) – environmental quality indicators related to urban environmental conditions, proxied through pollution-related measures.
- Social (S) – indicators reflecting safety and social well-being within the city environment.
- Governance (G) – institutional quality indicators reflecting governance effectiveness and administrative performance.

To maintain comparability across cities, the study uses indicators that are consistently available across all observations. All variables were collected for the closest available time period.

Environmental data were obtained from the Numbeo Pollution Index (2026 edition), which provides comparative measures of urban environmental quality across cities. Social indicators were derived from the Numbeo Safety Index, reflecting perceived safety and social conditions. Governance indicators were obtained from the World Bank Worldwide Governance Indicators (WGI), specifically the government effectiveness measure, which captures institutional quality and administrative performance.

2.2.3 Platform reputation indicators and data sources

Online reputation is measured using data extracted from TripAdvisor, one of the largest global digital tourism platforms. City-level reputation is operationalized through attraction-level user-generated data aggregated to the city level.

For each city, the data collection procedure follows a standardized process:

1. Open the city page on the selected tourism platform.
2. Navigate to “Things to Do” → “Attractions.”
3. Apply the same sorting rule for all cities.
4. Select the top 10 attractions for each city.
5. Record for each attraction:
   - the average rating scores
   - the total number of reviews.

The selection of the top 10 attractions for each city ensures consistency and comparability across observations, allowing reputation measures to be constructed using a standardized subset of highly visible tourist sites. These attraction-level observations are aggregated into city-level indicators representing an overall destination reputation on the platform.

The data were collected from TripAdvisor using the Traveler Favorites ranking filter on 15 March 2026.

2.3 Variable Construction

2.3.1 ESG measures

To enable comparability across cities, the environmental, social, and governance indicators are standardized before aggregation. Standardization allows variables measured on different scales to be combined into a unified sustainability index.

For each city i, the study constructs the following indicators:

- Environmental score: E_i
- Social score: S_i
- Governance score: G_i

An overall sustainability index is calculated as the arithmetic mean of the three dimensions:

$$ESG_i = (E_i + S_i + G_i) / 3 \qquad (1)$$

This composite indicator represents the overall sustainability performance of each city and serves as the primary explanatory variable in the empirical analysis.

2.3.2 Reputation measures

Using attraction-level ratings and review counts for each city, the study constructs two city-level indicators capturing different dimensions of digital tourism reputation.

Mean attraction rating (evaluative signal)

This indicator measures the average perceived quality of the city's most prominent attractions:

$$RatingMean\_i = (1 / K) \times \Sigma\ (j=1 \text{ to } K)\ r\_ij \qquad (2)$$

where K = 10 represents the number of attractions considered for each city and r_ij denotes the rating of attraction j in city i.

Log-transformed review volume (visibility signal)

Review counts are used as a proxy for platform attention and digital visibility. First, the total number of reviews across the selected attractions is calculated:

$$ReviewTotal\_i = \Sigma\ (j=1 \text{ to } K)\ c\_ij \qquad (3)$$

where c_ij represents the number of reviews for attraction j in city i.

Because review counts are typically highly skewed across destinations, a logarithmic transformation is applied:

$$ReviewLog\_i = \ln(1 + ReviewTotal\_i) \qquad (4)$$

This transformation reduces the influence of extremely popular destinations and improves comparability across cities.

2.4 Empirical Strategy

The empirical analysis is designed as a pilot quantitative test of the relationship between sustainability performance and digital tourism reputation. The objective is to identify preliminary statistical associations between ESG indicators and platform-based reputation signals.

The empirical strategy proceeds in three stages.

2.4.1 Descriptive analysis

The analysis begins with descriptive statistics for all variables in the dataset. Mean values, standard deviations, and minimum and maximum values are calculated to summarize variation across the sampled cities and identify potential outliers.

Table 1 presents the descriptive statistics for all variables.

2.4.2 Correlation analysis

The second step examines pairwise relationships between sustainability indicators and reputation measures. Correlation coefficients are calculated between:

- the composite ESG index and reputation indicators
- each ESG dimension (E, S, and G) and reputation indicators

This step provides initial evidence on whether sustainability performance is associated with digital reputation signals.

Table 2 reports the correlation matrix between ESG indicators and reputation variables.

2.4.3 Regression analysis

To examine these relationships more formally, the study estimates linear regression models with city-level reputation indicators as dependent variables.

First, the analysis evaluates the association between the composite ESG index and destination reputation:

$$Rep\_i = \alpha + \beta 1\ ESG\_i + \varepsilon\_i \qquad (5)$$

Second, the analysis decomposes sustainability into its individual components:

$$Rep\_i = \alpha + \beta E\ E\_i + \beta S\ S\_i + \beta G\ G\_i + \varepsilon\_i \quad (6)$$

where the dependent variable Rep_i represents either:

- RatingMean_i (average attraction rating), or
- ReviewLog_i (log-transformed review volume)

These models allow the study to assess whether sustainability performance and its individual dimensions are associated with variation in digital tourism reputation across cities.

The regression results are reported in Table 3.

To ensure methodological clarity, the analysis combines descriptive statistics, correlation analysis, and linear regression techniques. Correlation analysis is applied to assess the direction and strength of associations between ESG indicators and reputation measures. Linear regression models are then used to estimate the magnitude of these relationships and to examine how overall ESG performance and its individual components (E, S, G) contribute to variation in digital tourism reputation across cities.

Given the exploratory nature of the study and the limited sample size, the analysis is intended to provide preliminary evidence rather than generalizable conclusions.

3. Results and discussion

3.1 Descriptive statistics

Table 1 presents descriptive statistics for the variables used in the empirical analysis. The average attraction rating across the sampled cities is relatively high (mean = 4.50), indicating generally positive evaluations of major tourism attractions on digital platforms. Variation in average ratings is limited (SD = 0.14), suggesting relatively homogeneous evaluations across destinations.

In contrast, review volume shows greater dispersion. The logarithm of total reviews has a mean of 11.18 and ranges from 8.22 to 13.23, indicating substantial differences in platform visibility among cities.

Regarding sustainability indicators, environmental performance (E) exhibits the largest variability across the sample (mean = −61.91; SD = 17.34), reflecting differences in environmental conditions such as pollution levels. Social indicators (S), represented by safety-related metrics, display moderately high values (mean = 62.47; SD = 13.58). Governance indicators (G), based on institutional effectiveness measures, show a mean of 60.43 with comparatively lower variability (SD = 11.45).

The composite ESG index, calculated as the mean of the three dimensions, has an average value of 20.33 and ranges from 3.03 to 45.10. Overall, the descriptive statistics reveal substantial variation in sustainability indicators across cities, while reputation indicators, particularly average ratings, remain relatively stable.

Table 1. *Descriptive statistics*

Variable	Mean	Std. Dev.	Min	Max
rating_mean	4.50	0.14	4.09	4.68
review_log	11.18	1.64	8.22	13.23
E	-61.91	17.34	-90.56	-15.89
S	62.47	13.58	42	83.88
G	60.43	11.45	44.26	80.79
ESG	20.33	10.24	3.03	45.10

3.2 Correlation results

Table 2 reports pairwise correlation coefficients between reputation indicators and ESG-related variables. The correlation between average attraction ratings and review volume is weak (r = 0.041), indicating that higher ratings are not necessarily associated with greater platform visibility.

Environmental indicators show a moderate positive correlation with average attraction ratings (r = 0.357), while social indicators exhibit a smaller positive relationship (r = 0.181). Governance indicators display a relatively weak correlation with attraction ratings (r = 0.106).

In contrast, review volume demonstrates a strong positive correlation with governance performance (r = 0.669), suggesting that cities with stronger governance conditions tend to exhibit greater visibility on tourism platforms. The relationship between review volume and environmental indicators is weaker (r = 0.222), while social indicators display a negative correlation (r = −0.500).

As expected, the ESG composite index is strongly correlated with its individual components—particularly environmental performance (r = 0.880), followed by governance (r = 0.687) and social indicators (r = 0.558). These relationships confirm the internal consistency of the composite ESG measure.

Overall, the correlation results suggest that different sustainability dimensions may be associated with digital tourism reputation signals in different ways.

Table 2. *Correlation matrix*

Variable	rating_mean	review_log	E	S	G	ESG
rating_mean	1					
review_log	0.041	1				
E	0.357	0.222	1			
S	0.181	-0.500	0.225	1		
G	0.106	0.669	0.580	-0.029	1	
ESG	0.321	0.153	0.880	0.558	0.687	1

3.3 Regression results

Table 3 summarizes the regression estimates for the relationship between ESG performance and digital reputation indicators.

Model 1 examines the relationship between overall ESG performance and destination reputation measured by average attraction ratings. The results indicate a small but positive association (β = 0.0044), suggesting that higher sustainability performance is weakly related to higher online reputation scores. The explanatory power of the model is limited (R^2 = 0.103), which is expected given the exploratory nature of the pilot dataset.

Model 2 evaluates the relationship between ESG performance and platform visibility, measured by the logarithm of total review volume. The results show only a minimal association (β = 0.0246), with ESG explaining approximately 2.4% of the variation in review volume (R^2 = 0.024).

Models 3 and 4 disaggregate ESG into its three dimensions—Environmental (E), Social (S), and Governance (G). In Model 3, where average attraction ratings serve as the dependent variable, environmental performance shows the strongest positive association with ratings (β = 0.00335), followed by a smaller positive effect for social indicators (β = 0.00086). Governance indicators display a weak negative relationship (β = −0.00161). The model explains approximately 14.9% of the variation in attraction ratings (R^2 = 0.149).

Model 4 examines the relationship between ESG dimensions and platform visibility. Governance performance shows the strongest positive association with review volume (β = 0.10088), while environmental (β = −0.00791) and social (β = −0.05559) indicators exhibit small negative relationships.

The model explains a larger share of variation in visibility outcomes (R^2 = 0.683), indicating that governance conditions may be more strongly associated with tourism visibility than other sustainability indicators. This finding indicates that institutional quality and governance conditions may play a more significant role in shaping tourism visibility than environmental or social factors within platform-mediated environments. This pattern further suggests that digital platform visibility may be more closely linked to institutional and governance-related factors than to environmental or social conditions.

Given the limited sample size and exploratory design of the study, these findings should be interpreted as indicative rather than definitive.

Table 3. *Regression results*

Model	Dependent variable	ESG	E	S	G	R²
1	rating_mean	0.0044				0.103
2	review_log	0.0246				0.024
3	rating_mean		0.00335	0.00086	-0.00161	0.149
4	review_log		-0.00791	-0.05559	0.10088	0.683

3.4 Interpretation relative to hypotheses

Based on the empirical results, the hypotheses can be interpreted as follows.

- H1 (overall association): partially supported. The regression results show a small positive relationship between the composite ESG indicator and average attraction ratings, although the explanatory power of the model remains limited.
- H2 (dimension heterogeneity): Supported. When ESG is decomposed into its individual components, environmental and social indicators display small positive associations with attraction ratings, while governance indicators show a weak negative relationship.
- H3 (signal-type heterogeneity): Supported. Sustainability indicators relate differently to evaluative and visibility-based reputation signals. Environmental and social indicators appear more closely associated with perceived destination quality, while governance indicators exhibit a stronger relationship with platform visibility.

These results provide the basis for discussing the broader implications of ESG indicators for digital tourism reputation systems. This study adopts a pilot design with a limited sample size to provide initial exploratory evidence. The findings are not intended to be generalizable but rather to establish a preliminary analytical framework for future large-scale research.

4. Conclusion

This study examined whether sustainability performance, measured through Environmental, Social, and Governance (ESG) indicators, is reflected in digital reputation signals on tourism platforms at the city level. Using a pilot dataset of fifteen international tourism destinations, the analysis explored the relationship between ESG indicators and platform-based reputation measures derived from attraction ratings and review volumes.

The results suggest that the relationship between sustainability performance and digital tourism reputation is complex and heterogeneous. Environmental and social indicators appear to show modest positive associations with perceived destination quality, while governance indicators display a stronger relationship with platform visibility measured through review volume. At the same time, the explanatory power of the models remains limited, indicating that digital reputation on tourism platforms is influenced by multiple factors beyond sustainability performance alone.

These findings provide preliminary empirical evidence that sustainability-related conditions may influence different dimensions of digital tourism reputation. More broadly, the results highlight the potential relevance of integrating sustainability indicators into platform-based reputation systems in order to better align destination visibility with sustainable tourism objectives.

Given the limited sample size and exploratory design of this study, future research should extend the analysis using larger datasets and additional tourism platforms in order to further examine the relationship between sustainability performance and digital destination reputation.

Acknowledgment
Not applicable.

Authors' Declaration
Conflicts of Interest: The author declares no conflict of interest.
Ethical statement: This study used publicly available secondary data obtained from online platforms and international databases. No personal or sensitive user data were collected or processed. Therefore, ethical approval was not required.

Authors' Contribution Statement
Nijat Safarov conceptualized the study, collected and analyzed the data, developed the methodology, and prepared the manuscript.

References

1. Xiang, Z., Gretzel, U.: Role of social media in online travel information search. Tourism Management 31(2), 179–188 (2009). https://doi.org/10.1016/j.tourman.2009.02.016
2. Filieri, R.: What makes an online consumer review trustworthy? Annals of Tourism Research 58, 46–64 (2016). https://doi.org/10.1016/j.annals.2015.12.019
3. Litvin, S.W., Goldsmith, R.E., Pan, B.: Electronic word-of-mouth in hospitality and tourism management. Tourism Management 29(3), 458–468 (2007). https://doi.org/10.1016/j.tourman.2007.05.011
4. Srnicek, N.: Value, rent and platform capitalism. Edward Elgar Publishing (2021). https://doi.org/10.4337/9781802205138.00009
5. World Tourism Organization (UNWTO): UNWTO Tourism Highlights: 2018 Edition. UNWTO (2018). https://doi.org/10.18111/9789284419876
6. OECD: OECD Tourism Trends and Policies 2020. OECD Publishing (2020). https://doi.org/10.1787/6b47b985-en
7. Gössling, S., Hall, C.M., Weaver, D.: Sustainable tourism futures. Routledge (2009). https://doi.org/10.4324/9780203884256
8. Buhalis, D., Law, R.: Progress in information technology and tourism management: 20 years on and 10 years after the Internet—The state of eTourism research. Tourism Management 29(4), 609–623 (2008). https://doi.org/10.1016/j.tourman.2008.01.005
9. Poell, T., Nieborg, D., Van Dijck, J.: Platformisation. Internet Policy Review 8(4) (2019). https://doi.org/10.14763/2019.4.1425
10. Zheng, Y., Luo, Q., Wang, J.: Becoming famous: Producing platform-mediated destination visibility. Annals of Tourism Research 117, 104136 (2026). https://doi.org/10.1016/j.annals.2026.104136
11. Jeacle, I., Carter, C.: In TripAdvisor we trust: Rankings, calculative regimes and abstract systems. Accounting, Organizations and Society 36(4–5), 293–309 (2011). https://doi.org/10.1016/j.aos.2011.04.002
12. Ghose, A., Ipeirotis, P.G., Li, B.: Designing ranking systems for hotels on travel search engines by mining user-generated and crowdsourced content. Marketing Science 31(3), 493–520 (2012). https://doi.org/10.1287/mksc.1110.0700

13. Sparks, B.A., Browning, V.: The impact of online reviews on hotel booking intentions and perception of trust. Tourism Management 32(6), 1310–1323 (2011). https://doi.org/10.1016/j.tourman.2010.12.011
14. Zelenka, J., Azubuike, T., Pásková, M.: Trust model for online reviews of tourism services and evaluation of destinations. Administrative Sciences 11(2), 34 (2021). https://doi.org/10.3390/admsci11020034
15. Otero, J.M.M.: Fake reviews on online platforms: Perspectives from the US, UK and EU legislation. SN Social Sciences 1(7), 181 (2021). https://doi.org/10.1007/s43545-021-00193-8
16. Shukla, A.D., Goh, J.M.: Fighting fake reviews: Authenticated anonymous reviews using identity verification. Business Horizons 67(1), 71–81 (2023). https://doi.org/10.1016/j.bushor.2023.08.002
17. Manoli, A.E.: Platform-mediated local power in tourism. Annals of Tourism Research 118, 104151 (2026). https://doi.org/10.1016/j.annals.2026.104151
18. Torres-Delgado, A., Saarinen, J.: Using indicators to assess sustainable tourism development: A review. Tourism Geographies 16(1), 31–47 (2013). https://doi.org/10.1080/14616688.2013.867530
19. Font, X., Torres-Delgado, A., Crabolu, G., Martinez, J.P., Kantenbacher, J., Miller, G.: The impact of sustainable tourism indicators on destination competitiveness: The European Tourism Indicator System. Journal of Sustainable Tourism 31(7), 1608–1630 (2021). https://doi.org/10.1080/09669582.2021.1910281
20. Back, K.: ESG for the hospitality and tourism research: Essential demanded research area for all. Tourism Management 105, 104954 (2024). https://doi.org/10.1016/j.tourman.2024.104954
21. Lin, M.S., Zhang, H., Luo, Y., Li, Y.: Environmental, social, and governance (ESG) measurement in the tourism and hospitality industry: Views from a developing country. Journal of Travel and Tourism Marketing 41(1), 154–168 (2024). https://doi.org/10.1080/10548408.2023.2293008
22. Camilleri, M.A.: Environmental, social and governance (ESG) factors for sustainable tourism development: The way forward toward destination resilience and growth. Business Strategy and the Environment 35(3), 4057–4082 (2025). https://doi.org/10.1002/bse.70366
23. Shin, H., Jo, Y., Lee, S., Park, J., Cho, J., Kim, H.: ESG for sustainability in hospitality and tourism: A theoretical and practical review with a future research agenda. Journal of Travel Research 65(1), 3–36 (2025). https://doi.org/10.1177/00472875251363629
24. Arzoumanidis, I., Petti, L., Raggi, A.: Online booking platforms: Towards making more sustainable choices. Cleaner Production Letters 3, 100009 (2022). https://doi.org/10.1016/j.clpl.2022.100009
25. Hoffmann, F.J., Braesemann, F., Teubner, T.: Measuring sustainable tourism with online platform data. EPJ Data Science 11(1), 41 (2022). https://doi.org/10.1140/epjds/s13688-022-00354-6
26. Hajizade, S., & Balayev, R. Working methodology for assessing digital transformation risks. UNEC Journal of Computer Science and Digital Technologies
27. Gulmammadov, R. Digital Development Analysis based on cognitive modeling. UNEC Journal of Computer Science and Digital Technologies

Data Availability

The dataset used in this study was constructed from publicly available sources, including TripAdvisor attraction ratings and international sustainability indicators. The compiled dataset used in this analysis is available at:
[Google Sheets link]

UNEC Journal of Computer Science and Digital Technologies
Volume 2, № 1, 2026, pp. 76-99
https://doi.org/10.30546/UNECCSDT.2026.001.263

Making Inference Infrastructure Public: A Three-Layer Production Model for Small and Mid-Sized States

Oktay Ibrahimov
TF Consultancy, Baku, Azerbaijan
0009-0004-4033-1952, ogtay.ibrahim@outlook.com

Abstract

As AI evolves from an applied technology into the foundational substrate of economic coordination and public governance, the central strategic imperative for states has shifted: from ownership of frontier models to reliable, governable access to large-scale inference capacity. This paper extends the AI as Public Infrastructure framework by introducing Inference Infrastructure (I&I)—the nationally embedded capacity to generate and operationalize machine-mediated reasoning at scale—and argues it emerges only through synchronized co-development of three interdependent layers: physical compute and data infrastructure, operational AI systems capability, and institutional demand architecture. The absence of any single layer produces systemic underperformance, formalized through the multiplicative capacity function $I = C \times S \times D$. Integrating this I&I model with the Infrastructure Status Index, the paper introduces the *asynchrony penalty* — the systematic loss of inference capacity that follows from uneven development across the C, S, D layers — as the structural explanation for the documented gap between national AI strategy ambition and observable AI integration across the developing and middle-income world. The paper offers a diagnostic and policy framework for small and mid-sized states navigating the structural tension between digital sovereignty and integration with the global AI ecosystem. Drawing on evidence from the global shift toward inference-dominant AI workloads and national AI strategies—including Azerbaijan's AI Strategy 2025–2028 and Digital Economy Development Strategy 2026–2029, we demonstrate that the strategic objective for non-frontier states is not technological prestige but institutionalized access to machine reasoning capacity under conditions of governed interdependence.

***Keywords*:** AI governance, Inference Infrastructure, Public Infrastructure, Digital Sovereignty, Inference Economy, Governed Interdependence

Received:	Revised:	Accepted:	Published:
29/05/2026	03/06/2026	06/06/2026	17/06/2026

1. Introduction: From AI Adoption to Inference Capacity

1.1 The Fragmentation of AI Policy Discourse

National AI strategies — across advanced and developing economies alike — converge on a familiar repertoire: frontier model development, research ecosystem cultivation, startup acceleration, regulatory framework construction, and talent pipeline investment. Each element contributes to AI capability in isolation. Their disaggregated pursuit produces what we term *strategic scatter*: resource allocation across AI-adjacent objectives without a specified theory of how they combine to produce systemic economic output. In its absence, mid-sized and developing states default to policy mimicry: adoption of frontier-state strategic forms (ethics frameworks, talent pipelines, startup ecosystems, regulatory bodies) without the capital base, institutional density, or absorptive capacity that give those forms functional content. A special case of institutional isomorphism (DiMaggio and Powell, 1983), AI-policy mimicry is reinforced by the standardization of strategy templates circulated through multilateral organizations, consultancies, and development banks—producing global convergence in AI strategy form that masks systematic divergence in substance.

The persistence of mimicry despite documented failure rates—McKinsey (2025) estimates that 70–80% of agentic AI initiatives fail to scale beyond pilot, and the Federal Reserve Board (2025) documents systematic gaps between strategic commitment and realized AI integration—signals a structural constraint embedded in the current institutional environment. Correcting this requires a shift in the strategic object: from aggregated AI capability to inference capacity understood as a production output.

1.2 The Inference Turn

A structural shift in AI workload composition has reframed from the strategic question facing states. As AI mediates administrative decisions, supply chain optimization, energy management, legal drafting, financial modeling, and public service delivery, the relevant macroeconomic question is no longer whether a country develops AI but whether it can sustain domestically governable inference capacity embedded in institutional workflows.

Three terms carry analytical weight throughout. *Inference capacity* denotes the volume of machine-mediated reasoning operations a national system produces per unit time, capturing realized (operationalized) reasoning rather than latent capability, and remaining analytically distinct from model ownership and aggregate compute resources. *Domestic governability* is the subset of inference operations over which a state retains effective oversight, audit access, and continuity guarantees, resilient to vendor discretion, geopolitical realignment, or supply-chain disruption—positioned between territorial sovereignty (operations need not be domestically hosted) and mere commercial access. Workflow embedding is the structural integration of inference outputs into institutional decision processes such that their removal would degrade operational performance — the productive condition that distinguishes *inference infrastructure (I&I)* from *inference services*.

Non-frontier states, used throughout, denotes states that lack the capital scale, dataset access, and engineering depth to sustain frontier model pre-training, and whose strategic AI question is therefore deployment and governance rather than model ownership — a category that includes nearly all states outside the US and China and the conceptual focus of this paper. The category is internally heterogeneous along resource endowment, talent base, and geopolitical positioning — variation section 7 organizes through five trajectory profiles.

The empirical case for treating **inference capacity** and **domestic governability** as the relevant strategic variables rests on a documented inversion in the composition of AI compute. In 2023, model training accounted for approximately two-thirds of global AI compute, with inference comprising the remaining one-third. By 2025, this ratio had reversed, with inference projected to further expand its dominance through 2026 and beyond (Deloitte, 2025). This shift repositions the strategic locus of AI competition from capability formation (model training) to capability deployment (inference). Industry forecasts converge on inference market growth from approximately $113 billion in 2025 to over $250 billion by decade's end, with inference-optimized chip sales projected to exceed $50 billion in 2026 and AI data center capital expenditure reaching $400–450 billion globally (Deloitte, 2025). Power availability—defined as the provision of reliable, scalable, and cost-efficient electricity—has become the binding constraint, superseding raw compute capacity, as efficiency metrics shift from peak FLOPS to “tokens per watt per dollar.”

The structural logic is straightforward: frontier model training is finite and periodic; inference is continuous, scaling with every API call and embedded workflow. The strategic implication is consequential: under model training dominance, competition centers on model ownership—shaped by capital scale, dataset access, and frontier engineering talent. Under inference of dominance, the binding constraint shifts to deployment, integration, and governance capacity—competencies more evenly distributed and accessible to states that cannot sustain frontier model investment.

AI has entered its infrastructural phase not because models have grown larger but because inference outputs increasingly function as operational inputs into governance and economic coordination — a transition from AI as a tool to AI as a systemic substrate. Within the AIPI framework (Ibrahimov, 2025a), this corresponds to the movement toward public infrastructure status, requiring conceptual and measurement instruments adequate to its material substrate.

1.3 Scope and Contribution

This paper develops a framework for *inference infrastructure* (I&I) — the durable national capacity to generate, govern, and integrate machine-mediated reasoning into economic and administrative processes — and provides instruments for its diagnostic assessment. This contribution is fourfold.

First, the paper introduces I&I as a distinct analytical category, irreducible to compute, operational capability, or institutional demand, and emerging only through their synchronized development—relocating the strategic variable for non-frontier states from model ownership to operationalized inference under governable access.

Second, the paper specifies the multiplicative condition under which inference capacity is generated, captured heuristically as $I = C \times S \times D$. From it we derive the asynchrony penalty: the systematic loss of inference capacity from uneven layer development, and the structural explanation for the gap between AI strategy ambition and observable integration.

Third, the paper extends the Infrastructure Status Index (ISI) (Ibrahimov, 2025a) with inference-specific indicators that operationalize each of the three layers. These indicators are designed for diagnostic use: a state can identify which layer constitutes its binding constraint and adjust strategy accordingly. The extension converts AIPI from a conceptual framework into a measurable policy instrument.

Fourth, the paper applies the framework through illustrative comparative mapping across five small and mid-sized state (SMS) trajectories — Estonia, Singapore, the United Arab Emirates, Indonesia, and Azerbaijan. The cases extend the comparative mapping developed in companion work (Ibrahimov, 2026) by recoding each through the three-layer lens, and are selected to vary along the (C, S, D) profile. They are heuristic exemplars rather than empirical tests, consistent with the methodological orientation of the broader research program. Section 7 develops each case; the binding-constraint identification and strategic prescriptions follow in sections 9 and 10.

2. Theoretical Foundations and Literature Review

The framework developed in Section 4 sits at the intersection of four literatures, none of which individually provides an adequate account of inference as a productive variable: general purpose technology (GPT) theory, infrastructure economics, the political economy of compute, and digital sovereignty. We review each in turn, identifying both the analytical resources we draw on and the specific gap that motivates the present model. Jointly, these literatures establish that inference is consequential, that its provision exhibits public-goods characteristics, that its supply chain is structurally concentrated, and that sovereignty over it is contested — but none specifies the productive interaction among the inputs that determine whether a state can sustain it.

2.1 AI as General-Purpose Technology — and Its Analytical Limit

The classification of AI as a General-Purpose Technology (GPT) in the tradition of Bresnahan and Trajtenberg (1995) provides the standard framework for understanding its economic significance. Like electricity and digital computing before it, AI exhibits the three GPT hallmarks: pervasiveness across sectors, sustained potential for technical improvement, and the capacity to spawn complementary innovations. While the GPT framework captures the broad economic significance of AI and places it within a familiar lineage of general-purpose technologies, it is analytically insufficient for the present purpose. By treating technology as exogenous and centering diffusion dynamics—adoption lags, complementary innovation, and sectoral adjustment—it reduces institutional embedding to a downstream outcome rather than a constitutive element of production. In doing so, it abstracts away from the mechanisms and institutional logic through which states and organization's structure, govern, and sustain deployment at scale. Consequently, the framework remains confined to stylize historical comparison and specifies neither a production function for deployment within national institutional systems nor operational instruments for diagnosing why deployment fails.

A more recent line of work begins to address these limitations. Ide and Talamas (2025) provide the closest analytical precursor to the framework developed here. They model AI in the knowledge economy by treating inference compute—measured in FLOPS or token throughput—as a critical production input whose economic effects are mediated by organizational structure. Their model yields two results we adopt: first, that compute functions as a general-purpose productive resource; and second, that its impact depends not only on its quantity but on the organizational architecture through which it is deployed.

The present paper extends this framework into two dimensions. First, institutional architecture is treated as a coequal argument rather than a mediating variable. Second, operational capability is introduced as a distinct third argument, whereas Ide and Talamas subsume it within their organizational variable. This distinction is necessary because operational capability and demand architecture exhibit systematically different lead times, mobility properties, and failure modes—differences that Section 5 shows to be policy-relevant.

2.2 Infrastructure Economics and the Goods Classification Problem

Classical infrastructure economics classifies goods along the dimensions of excludability and rivalry, distinguishing private, club, and public goods. AI inference occupies a distinctive position within this framework. Raw compute is rival and excludable and thus constitutes a private good. Trained model weights are non-rival and non-excludable, though often rendered excludable through licensing and access controls, approximating a public good. Inference outputs embedded in institutional workflows are rival in capacity at the point of delivery yet generate non-rival positive externalities through productivity improvements that diffuse across sectors. Inference and infrastructure—the layered system enabling inference at institutional scale—is therefore best understood as an impure public good with congestion effects, analogous to electrical grids and telecommunications networks.

This classification has direct analytical consequences. Impure public goods with congestion are characterized by a structural tendency toward private underinvestment. Because private actors internalize only a portion of the social value generated, capacity provision systematically falls short of the welfare-optimal level. This property underpins the standard infrastructure economics rationale for public investment, regulatory oversight, and standardized access regimes. Historical transitions in telecommunications, electricity, and digital broadband illustrate this dynamic, whereby privately developed technologies are progressively reorganized as publicly governed infrastructures (Frischmann, 2012; Star and Ruhleder, 1996).

Emerging policy practice supports this classification: the EU's AI Factories initiative treats AI compute as shared public infrastructure, and South Korea's 2025 commitment to deploy over 260,000 GPUs across sovereign clouds represents national infrastructure investment rather than procurement.

We accept the impure-public-good classification as established. The infrastructure economics literature specifies provision rules for systems whose impure-public-good character is already operative — but not the structural conditions under which a national system attains that character. A state may invest in computing and produce neither systemic externalities nor congestion-relevant capacity if the institutional and operational layers required to convert compute into systemic inference output are absent. Section 4 specifies the conditions under which inference inputs combine to produce a system whose outputs warrant infrastructure-economics treatment in the first place.

2.3 The Political Economy of Compute: Concentration as a Production Constraint

The hardware supply chain for AI exhibits extreme concentration along its critical segments. Industry analyses consistently estimate NVIDIA's share of the AI GPU market in the range of 90–95%, with TSMC fabricating the overwhelming majority of advanced AI chips at the process nodes capable of supporting frontier inference. Three hyperscale cloud providers — Amazon Web Services, Microsoft Azure, and Google Cloud — mediate compute access for the bulk of global AI users, with a fourth tier of specialized providers (Oracle, CoreWeave, Lambda) operating within the same hardware dependency. Brookings (2026) characterizes the resulting structure as a system of concentrated choke points spanning minerals, energy, hardware, networks, infrastructure, data, models, applications, and the cross-cutting enablers of talent and governance.

For the framework specified in Section 4, this concentration yields a direct analytical implication not formalized in the existing literature. The compute argument, C, cannot be measured as a simple quantity—GPU count, aggregate FLOPS, or theoretical throughput—but must be specified as capacity conditional on access reliability. The issue is not ownership of compute, but guaranteed continuity of access. A state may rely on external providers under normal conditions; however, when access is contingent on the political and commercial discretion of a small number of foreign firms, nominal capacity overstates the level of compute that can be relied upon under disruption. We formalize this as the distinction between *gross compute* and *governable compute*: the former denotes total technical capacity available, while the latter denotes the subset over which the state retains continuity guarantees against geopolitical realignment, vendor decisions, and supply-chain disruption. Throughout the remainder of the paper, C refers to governable compute; gross compute appears only as a measurement caveat.

Luitse (2024) provides the anchor: cloud platforms exercise infrastructural power—the capacity to set conditions of possibility for AI production through control over the underlying stack. Federal Reserve Board (2025) quantifies the asymmetry: cumulative private AI investment 2013–2024 reached ~\$470 billion in the US against \$50 billion (EU), \$28 billion (UK), and \$6 billion (Japan), with the US hosting roughly four thousand data centers as of 2024. These differentials determine the discount factor converting gross to governable compute. A national strategy targeting a GPU count without specifying access governance has not specified C—only a number whose productive content depends on factors outside the state's control.

2.4 Digital Sovereignty: From Choice Dimensions to Production Constraints

Digital sovereignty discourse has moved beyond the binary between autonomy maximalism and market procurement. Recent work accepts that full-stack AI sovereignty is structurally infeasible even for the US and China (Brookings, 2026; Tony Blair Institute, 2026; World Economic Forum, 2026); the question has shifted to what kind of sovereignty, over which dimensions, under what conditions. McKinsey (2025) distinguishes territorial, operational, technological, and legal sovereignty; Ibrahimov (2026) extends this through governed interdependence and the Governance Membrane. None specifies the feasibility envelope for a state's (C, S, D) profile imposes — the gap between sections 4 and 8 address.

2.5 Synthesis: The Joint Gap

The four literatures identify components but none specifies the productive interaction among them: GPT theory treats embedding as diffusion outcome; infrastructure economics specifies provision rules without the conditions for attaining impure-public-good character; political economy documents compute concentration without formalizing it as a productive constraint; sovereignty discourse specifies dimensions without the feasibility envelope. Sections 3–4 address these gaps jointly.

3. AI as Executable Epistemic Infrastructure

This section develops the conceptual apparatus required to specify the production function in section 4. We argue that AI inference, as it is currently being deployed within institutional systems, constitutes a distinctive class of productive input — *executable epistemic infrastructure* — whose properties cannot be captured by treating inference as either a technological capability or an information-processing service. The conceptual primitives developed here become the structural arguments of the production function: compute as the substrate of execution, operational capability as the conversion mechanism, and institutional demand as the productive locus.

3.1 The Mechanization of Inferential Labor

Successive techno-economic paradigms have mechanized progressively more cognitive labor: electrification mechanized physical labor, digital computing mechanized arithmetic labor, AI mechanizes inferential labor — the synthesis and production of structured judgments where operations cannot be specified as deterministic procedures. Bureaucracies and professional bodies have historically performed this labor through institutional hierarchies (Garicano and Rossi-Hansberg, 2015), bounded by headcount and coordination cost. AI removes this scaling limit at the technical level only; institutional and governance constraints are repositioned, not removed.

3.2 What Makes I&I Executable

Earlier epistemic infrastructures—libraries, encyclopedias, statistical archives, expert systems—provided consultative outputs: structured knowledge that human users queried, interpreted, and applied at human speeds. They provided inputs to institutional cognition; they did not perform it. AI inference, integrated into workflows, produces executable outputs—structured judgments at machine speed, in formats institutional procedures can ingest directly without human re-synthesis.

Three properties distinguish executable from consultative epistemic infrastructure: temporal coupling (decision-relevant timescales enabling operational rather than only deliberative integration); format compatibility (structured outputs—decisions, classifications, recommendations—machine systems and procedures ingest directly); and volumetric continuity (continuous generation rather than episodic consultation). These properties make inference non-substitutable once embedded: an institution reorganized around continuous, structured, low-latency inference cannot revert to consultative provision without operational degradation — the structural dependence that establishes the ISI essentiality dimension (Ibrahimov, 2025a) and triggers AIPI-grade governance obligations.

3.3 Three Conditions of Productive Inference

The mechanization of inferential labor establishes the technical possibility of executable epistemic infrastructure. It does not establish its productive realization. National systems exhibit substantial variance in whether nominal inference capability translates into productive inference output, and the variance cannot be explained by differences in technological access alone. We argue that productive inference requires joint satisfaction of three conditions, each of which corresponds to one structural argument of the production function specified in section 4.

Substrate condition. Inference requires computing capable of generating outputs at workflow-relevant temporal and volumetric scales. It is necessary but not sufficient: a state with extensive compute that cannot be governably accessed has satisfied the substrate condition only for the entities controlling that compute — the gross-vs-governable distinction (section 2.3).

Conversion conditions. Compute does not autonomously produce institutionally usable inference. Domain adaptation, fine-tuning, integration, monitoring, and pipeline maintenance require operational capability embedded in human and organizational capacity. Technical-to-institutional translation is a productive activity, not pass-through. Documented failure rates (McKinsey, 2025; Federal Reserve Board, 2025) confirm conversion as a substantial empirical challenge — failures of conversion, not compute access (section 5.2).

Demand conditions. Inference outputs not institutionally consumed do not constitute productive output. The locus is the workflow: regulations mandating AI-assisted procedures, procurement incentivizing inference-enabled delivery, administrative architectures integrating inference into decisions, and cumulative learning from sustained use. Without demand-side embedding, compute and conversion capacity produce experimental output, not infrastructure.

Each of these three conditions is necessary; none is individually sufficient. Their joint operation is what produces I&I, and the structural relationship among them — formalized in section 4 — is what determines whether a national system attains the productive scale at which infrastructure-grade analysis applies.

3.4 I&I: Inference Infrastructure as Productive Variable

The conceptual apparatus developed in this section supports the following definition.

Definition 4 (I&I). The durable national capacity to generate executable epistemic outputs at workflow-relevant temporal, format, and volumetric scales, conditional on the joint satisfaction of substrate, conversion, and demand conditions, under governance arrangements that sustain the legitimacy and continuity of institutional inference dependence.

This formalizes section 3.3's three conditions, treating governance as a continuity requirement rather than an external constraint. The conditions enter section 4 as production-function arguments; governance enters a feasibility constraint on the function range (section 8). Within the AIPI program (Ibrahimov, 2025a), I&I is the productive substrate of the transition AIPI describes — the layer at which "AI is becoming public infrastructure" acquires material content as measurable capacity.

4. The Three I&I Layers and the Multiplicative Condition

Section 3 established that productive inference depends on three jointly necessary conditions — substrate, conversion, and demand. This section translates them into the structural arguments of I&I: three layers whose joint development determines a state's productive inference capacity. We specify the multiplicative condition under which the layers interact and draw out its implications for strategic priority-setting. Following the AIPI/ISI convention (Ibrahimov, 2025a), the layers are specified at the country–sector level; national-level diagnosis aggregates over sectoral profiles that may differ substantially, a point we return in section 10.3.

4.1 The Three Layers

I&I rests on three structurally distinct layers, each corresponding to one of the productive conditions developed in section 3.

Layer I — Compute and Data Infrastructure (C). The physical substrate required for inference execution at workflow-relevant scale (see Section 3.3).

Layer II — Operational AI systems capability (S). The human and organizational capacity to convert compute into institutionally usable inference. This includes machine learning engineers, data engineers, AI systems integrators embedded within ministries, sectoral domain translators, and MLOps professionals — together with the organizational architecture that hosts them.

Layer III — Institutional demand architecture (D). The regulatory, procurement, and organizational structures create systematic demand for inference within institutional workflows. Procurement standards, sectoral mandates, regulatory clarity, data governance frameworks, public-sector training, and performance metrics together constitute this layer.

The substantive treatment of each layer — its components, strategic challenges, characteristic failure modes, and observable indicators — is developed in section 5. The remainder of this section addresses the *interaction* among the layers, which is the structural property that distinguishes I&I from any of its constituent inputs.

4.2 The Multiplicative Condition

The three layers interact multiplicatively rather than additively. We capture this interaction in the heuristic expression: $\mathbf{I = C \times S \times D}$,

Where I represent inference capacity, and C, S, D represent the development levels of the three layers. The expression is a diagnostic heuristic, not an estimable production function — its purpose is to discipline strategic thinking, not support econometric inference. The multiplicative form captures the structural claim that no layer is dispensable; bounded substitutability among layers is plausible and addressed in section 4.4. It yields three policy implications: (1) all three layers are necessary; (2) balanced development outperforms concentrated investment in any single layer; (3) diagnosis of the binding constraint must precede prioritization. These are applied in section 5 and section 7.

4.3 The Asynchrony Penalty

The multiplicative condition implies a phenomenon we term the **asynchrony penalty**: the systematic loss of inference capacity that follows from uneven layer development. A state that invests heavily in one or two layers while allowing the third to lag operates substantially below the inference capacity its total investment could otherwise sustain, even if its absolute spending matches that of states with balanced layer development. The asynchrony penalty is not a hypothetical loss; it is the structural explanation for the documented gap between AI strategy ambition and observable AI integration that motivates the present paper.

The empirical signature is distinctive: states exhibiting the penalty have credible strategies — substantial commitments, announced investments, visible flagship initiatives — yet operate in workflows where AI integration remains shallow. Compute is built but underutilized; capability is trained but exported; demand architecture is announced but not enforced. Reducing the penalty requires not more investment in the visible layers but closing the gap with the lagging one.

The asynchrony penalty also clarifies the structural mechanism behind the policy mimicry problem identified in section 1.1: states that adopt frontier-state strategic templates without the institutional density to execute them replicate the form of allocation across layers without the substance of any individual layer's development. The result is a high asynchrony penalty — a structural outcome the multiplicative condition predicts, and that observed strategy outcomes confirm.

4.4 What the Multiplicative Condition Does Not Imply

Three clarifications protect against overreach. First, the multiplicative condition is a structural claim about layer-interaction direction, not a mathematical identity: capacity does not equal the product of three indices, and the expression is not econometrically estimated. Second, layers need not develop simultaneously; sequencing is consistent with the condition provided lagging layers do not remain near zero for sustained periods. Third, bounded inter-layer substitution is plausible — operational capability can partially compensate for compute scarcity — but no layer is dispensable. Once any layer is being developed, the others must be paced against it.

5. The Three Layers in Detail

This section develops each of the three layers introduced in section 4 — components, strategic challenges, and characteristic failure modes. Observable indicators are operationalized in section 6; country-specific material is deferred to section 7.

5.1 Layer I: Compute and Data Infrastructure

Components. Layer I encompass the physical substrate required for inference execution: GPU and accelerator capacity, AI-optimized data centers, edge compute, and high-bandwidth networks (compute); secure national data environments with residency and access governance, structured datasets, and interoperability standards (data); and the power generation, transmission, and load-management capacity required to sustain AI workloads at scale (energy). The three sub-components are functionally interdependent: compute without data is idle, data without compute unprocessable, both without energy inert. Layer I must satisfy the temporal-coupling, format-compatibility, and volumetric-continuity conditions (section 3.2) without which compute does not translate into infrastructural substrate.

Strategic challenges. The principal challenge is not absolute capacity but governable capacity — the subset over which the state retains continuity guarantees against geopolitical realignment, vendor decisions, and supply-chain disruption. Most small and mid-sized states will depend on imported hardware and cloud supplements indefinitely; the objective is not to eliminate dependence but to construct it under terms of preserving governance access. Energy carries particular weight: Norris et al. (2025) find the existing U.S. grid could integrate roughly 100 GW of new AI load through 2029 if workloads were scheduled with modest flexibility — implying energy and AI strategy must be co-designed. For non-frontier states, the co-design imperative is more acute, since the margin for inefficient utilization is smaller.

Characteristic failure modes. Three patterns recur. *Stranded compute*: GPU clusters and data centers built but underutilized because Layers II and III cannot convert raw compute into productive output. *Sovereign exposure*: nominal compute capacity high but governable capacity low operations technically supported but politically contingent on external actors. *Energy bottlenecks*: compute built without coordinated energy provisioning, producing utilization caps and rising marginal costs. All three reflect investment in Layer I's visible components without attention to its less visible ones.

5.2 Layer II: Operational AI Systems Capability

Components. Layer II comprises the human and organizational capacity required to convert compute and data into productive inference output. Five role-types constitute its core: machine learning engineers (model adaptation, fine-tuning, domain optimization); data engineers (pipeline management and data-quality maintenance); AI systems integrators embedded within ministries and public agencies; sectoral domain translators with training in both AI methods and a substantive domain (medicine, law, public administration); and MLOps professionals maintaining stable, auditable inference operations in production. Beyond these roles, the layer encompasses organizational capacity to develop AI applications suited to national linguistic, legal, and administrative contexts — generic frontier models trained primarily on English-language data perform unevenly on local languages, legal systems, or administrative procedures.

Strategic challenges. Operational capability cannot be purchased off the shelf or developed quickly. Lead time from investment to productive deployment exceeds Layer I (compute is installed in months; engineers and integrators require years to train and embed). International labor markets compound the challenge: jurisdictions with deeper Layer III outcompete those without talent retention. A state that invests in capability without simultaneously developing demand produces an export commodity, not a national resource — the coordination problem the multiplicative condition makes explicit.

Characteristic failure modes. *Capability of export*: operationally capable professionals migrate to jurisdictions with better deployment opportunities, leaving training costs without productive capacity. *Pilot stagnation*: AI initiatives reach demonstration stage but fail to scale; industry evidence suggests 70–80% of agentic AI initiatives struggle beyond pilot, indicating a dominant rather than marginal outcome. *Vendor capture*: external cloud providers and consultancies absorb the operational function the state failed to develop domestically, with value flowing offshore and institutional learning lost. All three are invisible at the strategy-document level—a state can have a credible strategy, substantial compute, and visible initiatives while exhibiting all three simultaneously.

5.3 Layer III: Institutional Demand Architecture

Components. Layer III comprises the regulatory, procurement, and organizational structures that create systematic demand for inference within institutional workflows. Six components are observable: procurement standards requiring or incentivizing AI-enabled services in public contracting; sectoral programs mandating AI integration with explicit performance targets; regulatory clarity on data use that reduces the institutional risk premium on AI deployment; national data governance frameworks specifying conditions for using administrative data to train domain-specific AI; public-sector training that builds absorptive capacity; and performance metrics accounting for AI integration in evaluating institutional outcomes.

Strategic challenges. Demand for inference must be constructed; it does not arise spontaneously from compute and capability availability. Institutions exhibit predictable resistance: existing workflows are sunk-cost investments; professional identities are bound to non-AI processes; accountability structures are calibrated to non-AI outputs. Demand architecture is the set of institutional changes that make AI integration the path of least resistance. Timing is a second challenge: demand cannot precede capability (no supply to satisfy it) nor lag it (unmet capability exports). Demand and capability must be staged together.

Characteristic failure modes. *Strategy-without demand*: strategy documents, working groups, and budget commitments are announced while procurement, regulatory, and performance-metric structures remain unchanged—demand signaled but not constructed. *Sectoral isolation*: AI integration succeeds in one or two flagship sectors (often e-government) while remaining absent across the broader institutional landscape. *Regulatory paralysis*: uncertainty about data use, liability, and audit suppresses institutional appetite for AI integration even when other layers are developed, with institutions calculating the risk premium exceeds expected operational benefit and deferring indefinitely.

Assessment of the binding constraint across all three layers requires measurement instruments adequate to this structure, developed in the next section.

6. Extending the ISI: Measurement of I&I

The Infrastructure Status Index (ISI), developed within the AIPI framework (Ibrahimov, 2025a), measures the degree to which AI within a given country–sector pairing has crossed into public-infrastructure status along four dimensions: Essentiality, Embeddedness, Legitimacy, and Governance. ISI provides the diagnostic structure through which AIPI claims about infrastructural transition are converted into measurable assessments. This section extends ISI with a set of *inference-specific indicators* organized by the three-layer structure developed in section 4 and section 5. The extension allows ISI to function not only as a diagnostic of infrastructural status but as an instrument for identifying the binding-constraint layer and adjusting strategy accordingly.

6.1 The Logic of the Extension

The original ISI dimensions ask whether AI has crossed into infrastructure-grade significance. The indicators below answer the prior question: what is each layer's development level, and which is the binding constraint? A state crossing ISI threshold with a near-zero layer has reached infrastructural status under high asynchrony penalty — institutional dependence exceeding capacity to govern it. A state balanced across layers but not crossing ISI thresholds has built productive capacity without infrastructural significance; its challenge is embedding, not scaling. The indicators are designed for diagnostic use rather than cross-country ranking (section 10).

6.2 Layer I Indicators: Substrate Capacity

Three indicator concepts assess the substrate condition.

I-1: Domestic Inference Execution Ratio. The proportion of national inference workloads executed on domestically governable infrastructure (per section 2.3) — not raw compute installed but compute available under continuity-protected access conditions. *Proxies*: cloud billing data segmented by jurisdiction; data center utilization reports; government IT procurement records.

I-2: Sovereign Compute Diversification. The number and concentration of compute-supplier relationships available to domestic institutions, weighted by continuity-guarantee credibility. A state with a single hyperscale supplier scores lower than one with three diversified suppliers of comparable aggregate capacity. *Proxies*: vendor concentration ratios in public-sector procurement; multi-cloud arrangement enforceability; supply-chain audit records.

I-3: Energy-Compute Co-Adequacy. The proportion of allocated AI compute capacity sustainable under projected energy supply over a five-year horizon, accounting for grid flexibility and AI workload profiles. *Proxies*: power capacity allocated to AI workloads; grid flexibility assessments; forward energy adequacy projections.

6.3 Layer II Indicators: Conversion Capacity

Three indicator concepts assess the conversion conditions.

II-1: Inference Workforce Density. Per-capita's availability of the role-types specified in section 5.2, weighted by sectoral distribution and retention rates. Retention weighting is essential: density not retained in domestic deployment generates the capability-export failure mode. *Proxies*: labor force surveys; professional certification databases; tax-residence and employment-status data on certified professionals.

II-2: Institutional Embedding Capability. Presence and depth of AI systems integrators within ministries and major public agencies, and MLOps practice maturity in critical sectors. Capability institutionally co-located is productive in ways that capability located only in the private sector or research institutions is not. *Proxies*: organizational audits of public agencies; agency-level surveys on AI integration roles; production-system maturity assessments.

II-3: Domain Adaptation Capacity. Existence and scale of national-language model adaptation, sectoral fine-tuning, and locally produced AI applications in public-administration use — the ability to produce inference output adapted to national linguistic, legal, and administrative contexts rather than deploy generic frontier outputs as-is. *Proxies*: catalogues of nationally fine-tuned models; sectoral inventories of AI in production use; national-language NLP capacity documentation.

6.4 Layer III Indicators: Demand Architecture

Three indicator concepts assess the demand condition.

III-1: AI Procurement Penetration Rate. The share of public-sector procurement requires or incentivizing AI-enabled service delivery, weighted by contractual depth (preference, requirement, or core specification). Procurement converts institutional demand from signal into operational market. *Proxies*: procurement database analysis; contract reviews; tender specifications coded for AI requirements.

III-2: Workflow Integration Index. The degree to which AI inference outputs enter institutional decision processes as primary inputs (used directly in decisions) rather than supplementary inputs (advisory information re-synthesized by human decision-makers). Primary-input integration generates the structural dependence defining infrastructural status; supplementary integration does not. *Proxies*: agency surveys on AI use in administrative determinations; workflow analysis in priority sectors; documentation of AI outputs in regulatory and procurement decisions.

III-3: Regulatory and Governance Maturity. Comprehensiveness and enforcement strength of regulatory frameworks for AI integration — data governance, audit and transparency requirements, liability rules, and exit-rights provisions in contracts with foreign providers — the institutional structures that make AI integration viable for risk-averse public-sector institutions. *Proxies*: regulatory framework assessments; audit mechanism reviews; contract clause analysis.

6.5 Diagnostic Use

The extension supports diagnostic applications along three lines. *Identifying the binding-constraint layer*: the indicator profile reveals which of C, S, D is most depressed relative to peers or relative to the state's own development level in other layers — the prerequisite for the policy prescriptions in section 9. *Locating asynchrony*: distinct profiles (high-C/low-D, high-D/low-C, balanced) correspond to distinct strategic trajectories developed in section 7. *Tracking change over time*: repeated measurement reveals whether layer development is rebalancing toward the binding constraint or whether the asynchrony pattern is deepening.

Full operationalization — weighting schemes, sampling protocols, cross-jurisdictional comparability — is an open research program (section 10). The present contribution specifies the indicator's concepts and their structural relationship to the multiplicative condition; empirical implementation, including in-country pilots with statistical agencies, is a task for follow-up work.

7. Five Trajectories: Comparative Mapping of Layer Profiles

This section applies the three-layer model to five small and mid-sized state trajectories — Estonia, Singapore, the UAE, Indonesia, and Azerbaijan — recoded through the (C, S, D) lens to identify each case's binding-constraint layer and characteristic failure modes. The cases extend the comparative mapping developed in Ibrahimov (2026) and reuse its typology labels: Estonia as governance-led, Singapore as hub-positioned, the UAE as capital-led, Indonesia as scale-driven, and Azerbaijan as a transitional case with operational capability as the binding constraint. Azerbaijan’s profile is sketched here for comparative positioning; a detailed analysis appears in Section 10.

7.1 Estonia — Governance-Led Trajectory

Estonia represents the analytical limit of how far Layer III development can compensate for limited Layer I capacity. Its X-Road infrastructure provides one of the most institutionally embedded digital-government architectures globally, and the Kratt AI strategy (2019) extends this embedding into AI-mediated public administration. AI integration is mandated across e-government services, procurement standards specify AI-enabled service delivery, and operational redundancy with explicit service-level requirements acknowledges systemic dependency on inference outputs (Ibrahimov, 2025a). Layer II is moderately developed against population scale, with sectoral domain translators present in major regulated sectors (health, justice, finance) and mature MLOps practice in production e-government systems.

Layer I is the structural limit. Estonia does not maintain frontier-scale domestic compute and remains structurally reliant on foreign cloud infrastructure within the EU AI Factories arrangement and EU hyperscaler partnerships; governable compute is partially secured through EU-level governance but is not domestically constituted. The profile is high-D, moderate-S, low-C: productive contribution is bounded by the low C. As inference systems transition from administrative tools to cognitive infrastructure, governance strength alone may become insufficient where deeper integration with domestic compute is a productive requirement. Estonia's continued effectiveness depends on the stability of the EU compute environment — a dependence the framework recommends managing through diversified access within the EU AI Factories structure rather than treating as resolved by EU membership alone.

7.2 Singapore — Hub-Positioned State

Singapore exhibits the most balanced layer profile of the five cases and illustrates what coordinated three-layer development at small-state scale can achieve. Layer II is exceptionally developed by small-state standards, anchored by AI Singapore (AISG), the SEA-LION sovereign language model initiative, and a sustained pipeline of ML engineers and data scientists trained through university–industry partnerships. Capability retention is supported by deep institutional demand in finance, health, and public administration, with Singapore's hub role generating productive domestic deployment opportunities. Layer III is similarly mature: The Model AI Governance Framework (IMDA, 2024) institutionalizes transparency, accountability, and human-oversight requirements; procurement standards consistently incentivize AI-enabled service delivery; and AI integration has moved beyond flagship initiatives into routine operational use.

Layer I is moderate but consistent with Singapore's posture. The country does not maintain frontier-scale domestic compute and remains deeply integrated into global cloud infrastructures, but its compute relationships are diversified and contractually deep: hyperscaler arrangements are negotiated from positions of governance maturity, and operational governance over compute access exceeds what the raw footprint suggests. The profile is high-S, high-D, moderate-C, with moderate-C deliberately constructed under terms that approximate governable status. The multiplicative condition predicts strong productive output, and observed outcomes are consistent. The principal risk is that hub-positioned status depends on continued diplomatic and commercial standing in a multipolar AI landscape; the diversification logic that protects Singapore today may face stress under geopolitical bifurcation.

7.3 United Arab Emirates — Capital-Led Trajectory

The UAE provides the clearest illustration of how heavy investment in the most visible layer can produce inference capability without proportionally producing I&I. Layer I development has been aggressive and well-resourced: the 2025 US–UAE AI Acceleration Partnership granted advanced semiconductor access and a planned 5-gigawatt AI campus in exchange for explicit governance and security commitments (Ibrahimov, 2026); G42 and related sovereign-fund vehicles have built frontier-grade GPU clusters and AI-optimized data centers at a scale unusual among small states; and Falcon, the UAE-developed open-weight model series, demonstrated domestic capacity for model development. By raw compute metrics relative to population, the UAE ranks among the most-capitalized AI-infrastructure states globally.

Layer II is growing but uneven cultivated through international talent attraction and flagship initiatives (health diagnostics, energy, smart cities), with operational embedding within ministries and MLOps practice shallower than the compute footprint suggests and capability concentrated in flagship organizations. Layer III is the structurally constrained layer: integration is deepest in flagship sectors and citizen-facing services, while routine institutional demand remains thin. The UAE's earlier framing of frontier AI as a public service — free national GPT-4 access alongside domestic compute investment (Ibrahimov, 2025a) — is a creative response to the demand deficit but a substitute for institutional embedding, not its constructed equivalent.

The profile is high-C, growing-S, moderate-D. The multiplicative condition predicts diminishing returns to further compute investment relative to demand-side embedding; demand architecture, given its slower lead time, should not be deferred.

7.4 Indonesia — Scale-Driven Trajectory

Indonesia is the largest case in our comparison and the most complex layer profile. As an archipelago of more than 17,000 islands spanning several major linguistic communities, Indonesia operates under structural conditions that make uniform layer development inherently difficult; layer development is best read as uneven across geography and sector rather than monotonically high or low.

Layer I is in an active build-out. The flagship achievement is GPU Merdeka, the national GPU infrastructure that hosts Sahabat-AI, launched as an 8B–9B sovereign SLM family by Indosat Ooredoo Hutchison and GoTo and scaled to 70B-parameter variants on open-weight foundations, supporting Bahasa Indonesia and major regional languages — Javanese, Sundanese, Balinese, Bataknese — on national compute (Indosat & GoTo, 2024; 2025). Large-scale AI-ready data centers and a proposed Sovereign AI Fund signal substantial Layer I commitment, though AI-optimized capacity remains concentrated in major urban centers. Layer II is closely coupled with Sahabat-AI's deployment, cultivated through the Indosat–GoTo partnership and public-sector training aligned with the National AI Strategy (2020) and the 2025 AI National Roadmap. The choice of open-weight foundations over frontier model development is scale-appropriate: it allocates operational capability to adaptation — fine-tuning, multilingual extension, sector-specific deployment — rather than frontier engineering. Across an archipelagic institutional landscape, adaptation capability outproduces frontier capability per dollar invested.

Layer III is uneven: Sahabat-AI powers public services, education, and regulatory compliance under data-localization and Pancasila-rooted guidelines, but demand architecture is deep in flagship initiatives and urban institutions, thinner in peripheral provinces and across judicial, regulatory, and sectoral layers. The profile is moderate-C, moderate-S, uneven-D. For scale-driven trajectories, the binding constraint is consistency across the national territory rather than level of any single layer at the national average; extending demand architecture into lagging regions yields higher returns than further compute build-out in leading regions.

7.5 Azerbaijan — Transitional Case

Azerbaijan presents a transitional profile in which compute and demand architecture are emerging while operational capability constitutes the binding constraint. Layer I is in active development, with the TransCaspian Fiber Optic project, the 2025 launch of AzInTelecom's national high-performance computing centre, and diversified cloud partnerships with AWS, Google Cloud, and Microsoft Azure (Ibrahimov, 2026). Layer III shows emerging strength: the ASAN service architecture and over 450 e-government services provide an institutional substrate for AI demand, and the AI Strategy 2025–2028 (Republic of Azerbaijan, 2025a) and Digital Economy Development Strategy 2026–2029 (Republic of Azerbaijan, 2025c) signal genuine institutional commitment. Layer II is the binding constraint: the current base of AI operations professionals — ML engineers, data engineers, integrators, MLOps specialists — remains small relative to the strategy's ambitions, and brain drain to higher-paying jurisdictions is a significant risk.

The profile is moderate-C, low-S, moderate-D. The strategic implication, developed in section 10.3, is that operational capability should be prioritized over further compute investment in the near term, with demand architecture paced against capability build-up to avoid the capability-export failure mode.

7.6 Cross-Cutting Synthesis

Two patterns emerge. The binding-constraint layer varies by context while the structural logic holds: Estonia is constrained by Layer I, the UAE by Layer III, Azerbaijan by Layer II; Singapore approaches balance and Indonesia exhibit geographic unevenness rather than a single national constraint. Visibility differentials systematically bias investment toward the most photographable layer — the UAE illustrates this; Estonia the inverse. Binding-constraint diagnosis therefore requires explicit measurement (section 6).

8. Governed Interdependence as Strategic Configuration

The multiplicative condition (section 4) and comparative mapping (section 7) establish the structural conditions under which I&I can be produced. They do not specify the strategic posture toward the global AI ecosystem within which those conditions must be met. This section closes that gap: governed interdependence (Ibrahimov, 2026) is the configuration structurally implied by the multiplicative condition for non-frontier states — not the right posture because more sovereign or more cooperative than the alternatives, but the only one consistent with the production conditions the multiplicative condition specifies.

8.1 The Structural Exclusion of the Alternatives

Two strategic postures dominate contemporary AI sovereignty discourse for small and mid-sized states. *Autonomy maximalism* aims at full domestic control over the AI value chain — chip fabrication through model training to application deployment — under the assumption that strategic security requires technological self-sufficiency. *Uncritical integration* treats AI infrastructure as a procurement decision subject to standard commercial considerations, accepting whatever supplier relationships markets generate. The multiplicative condition rules both out, on structural grounds developed below.

Autonomy maximalism is structurally infeasible: the substrate condition requires hardware supply chains concentrated at NVIDIA, TSMC, and the major hyperscale's (section 2.3). No state below frontier scale has produced inference-grade compute domestically against these dependencies, and the recent acknowledgment that full-stack sovereignty is infeasible even for the US and China (Brookings, 2026; Tony Blair Institute, 2026) confirms the limit binds at the highest levels. For small and mid-sized states, maximalism diverts resources toward the least achievable layer, deepening the asynchrony penalty rather than reducing it.

Uncritical integration is structurally vulnerable because the productive contribution of inference depends on *governable* access conditions, not on access alone. The gross-vs-governable distinction developed in section 2.3 establishes that compute available under contractual terms that can be modified by foreign vendor decisions, geopolitical realignment, or supply-chain disruption is productively discounted relative to compute available under continuity-protected access conditions. A state that procures compute on standard commercial terms without specifying governance, audit, exit-rights, or diversification provisions has not specified C in the multiplicative condition; it has specified a quantity whose productive content is contingent on factors outside its control. Where these contingencies trigger — vendor decisions to terminate access, regulatory changes in the supplier jurisdiction, supply-chain interruptions — the productive value of integration collapses, and the state's institutional dependence on inference outputs becomes operationally vulnerable. In the language of section 3.2, uncritical integration produces *inference services* but not *inference infrastructure (I&I)*.

Both alternatives fail for the same reason: neither is consistent with sections 3–4's production conditions. Maximalism mis-allocates across layers; uncritical integration mis-specifies access conditions. The structural exclusion follows from the productive conditions of inference itself, not strategic preference.

8.2 Governed Interdependence: Layer-Specific Specifications

Governed interdependence is the strategic configuration in which a state participates in the global AI ecosystem under terms that preserve governance access across each productive layer. The configuration is not a single posture, but a layer-specific set of commitments derived from the three-layer model.

At Layer I. Governed interdependence requires compute access satisfying the gross-vs-governable distinction: diversified supplier relationships (no single hyperscaler accounting for capacity the state cannot afford to lose), contractual continuity guarantees (audit access, exit rights, advance-notice provisions), domestic anchor capacity sufficient for critical-function continuity under foreign-supplier disruption, and explicit data-residency and inference-pipeline jurisdiction. The configuration is not autarkic but structurally distinct from uncritical integration: governance properties are negotiated as primary terms rather than commercial residuals. Singapore illustrates this in mature form; Estonia in EU-mediated form; the UAE in transitional form, with quantity-vs-governability still weighted toward quantity.

At Layer II (Operational AI Systems Capability). Governed interdependence specifies that operational capability be cultivated domestically under conditions that retain it productively: workforce development calibrated to deployment rather than frontier engineering (the Indonesia open-weight adaptation approach is structurally appropriate), embedding of integrators within ministries rather than concentration in research institutions, and Layer III demand architecture sufficient to retain capability against international labor-market gravity. Layer II governance has no fully independent content — it is structurally inseparable from Layer III — because operational capability cannot be "governed" except through the demand of architecture that retains it. The capability-export failure mode (section 5.2) is therefore the structural risk against which the joint Layer II / Layer III specification must be designed.

At Layer III (Institutional Demand Architecture). Governed interdependence specifies that institutional demand be constructed under terms that preserve domestic governance over inference-mediated decisions: procurement standards that specify governance properties (audit access, explainability requirements, liability allocation) alongside performance properties; sectoral mandates with explicit accountability arrangements; regulatory frameworks that establish data governance as enabling rather than residual; and performance metrics that account for inference integration in evaluating institutional outcomes. The McKinsey (2025) four-dimensional framework discussed in section 2.4 operates at Layer III, specifying the dimensions across which sovereignty configurations vary; the present framework specifies which configurations are structurally available to states with which layer profiles.

8.3 The Governance Membrane as Operational Architecture

The strategic posture specified above requires institutional architecture adequate to its operationalization. The Governance Membrane (Ibrahimov, 2026) is developed in companion work as the reference architecture for institutionalizing governed interdependence. The present paper does not re-derive that architecture — its specification is the substantive contribution of the companion paper — but it is useful to indicate the connection between the production-function framework developed here and the governance architecture developed there.

The Governance Membrane is a layered institutional architecture — comprising oversight bodies, audit interfaces, contractual standards, and exit-rights provisions — that mediates the boundary between domestic AI systems and global AI infrastructure (Ibrahimov, 2026). It filters rather than blocks integration: data sovereignty requirements are enforced at defined interface points; inference operations affecting critical institutional functions are subject to specified domestic oversight; AI service terms are negotiated from positions of institutional agency rather than from structural dependency. The Normative Compliance Model, the Infrastructure Status Index, and the Cognitive Dependence Index operate within the membrane architecture, with ISI providing the diagnostic backbone connecting section 6's measurement framework to the operational governance architecture.

The section 4 framework and the Governance Membrane are jointly constitutive: production conditions without governance architecture produce ungovernable capacity; governance architecture without production conditions produces governance with no substrate. The task is simultaneous construction of both.

9. Policy Implications

The framework developed in sections 3–8 generates two classes of policy prescription: universal commitments that follow from the multiplicative condition and apply to all non-frontier states, and profile-specific prescriptions that follow from the binding-constraint layer identified through section 6's diagnostic apparatus and illustrated in section 7. We treat each in turn, then close with sequencing implications.

9.1 Structural Exclusions: What the Framework Rules Out

The multiplicative condition rules out two strategic postures on structural grounds, as developed in section 8.1. The corresponding policy implications are:

Avoid frontier model races. Pre-training a frontier large language model requires capital, dataset access, and engineering scale systematically inaccessible to non-frontier states — and crucially, frontier model development addresses none of the three layers that determine inference capacity: it does not provide governable compute, does not develop deployment-oriented capability (frontier engineering and deployment engineering are distinct skill sets), and does not construct institutional demand. The Indonesia approach — building atop open-weight foundations and reallocating capability to adaptation, fine-tuning, and multilingual extension — is structurally appropriate and worth generalizing.

Avoid uncritical procurement of inference services. Procurement on standard commercial terms without specifying governance properties produces inference services but not I&I (section 3.2). Procurement of inference-related services should specify the governance properties identified in section 8.2: contractual continuity guarantees, audit access, exit rights, advance-notice provisions, and explicit jurisdiction over data residency and inference-pipeline operation. The cost of negotiating these properties at contracting is substantially lower than retrofitting them after institutional dependence has developed.

9.2 Diagnose Before Investing

The most consequential universal prescription is *diagnostic primacy*: identification of the binding-constraint layer before allocation of new investment. Strategic priority cannot be assigned in absolute terms because the multiplicative condition implies that the layer most worthy of additional investment varies by national context. The diagnostic apparatus developed in section 6 — domestic inference execution ratio, sovereign compute diversification, energy-compute co-adequacy, inference workforce density, institutional embedding of capability, domain adaptation capacity, AI procurement penetration rate, workflow integration index, regulatory and governance maturity — provides the instrument through which diagnosis is conducted.

The policy implication is institutional. States that have built AI strategy infrastructure (national strategies, AI ministries, working groups) without diagnostic capacity make systematically biased investment decisions, allocating resources toward the most visible layer rather than the binding-constraint layer. The implied institutional reform is the development of AI strategy diagnostic units — cross-ministry teams with the analytical capacity to assess national layer profiles using section 6's indicators, and the institutional standing to recommend reallocation of investment based on diagnostic findings. The unit need not be large but must be analytically credible and procedurally consequential.

9.3 Design for Synchronization

The multiplicative condition implies that the three layers must be developed in coordination: investment in any one layer should be paced against the development levels of the other two to avoid the asynchrony penalty. Synchronization requires three institutional reforms.

First, AI strategy governance must span the boundaries conventionally separating compute infrastructure (digital/telecoms ministries), workforce development (education/labor), and institutional procurement (finance and sectoral ministries). Policy mimicry (section 1.1) is partly a failure of inter-ministerial coordination: each ministry pursues its remit, and the asynchrony penalty emerges from the absence of coordination across remits. Synchronization requires AI strategy bodies at the head-of-government level with mandate authority across ministries.

Second, budget cycles for the three layers must be coordinated. Compute investment cycles operate on three-to-five-year horizons (data center construction, hardware procurement).

Capability development cycles operate on five-to-ten-year horizons (workforce training, institutional embedding). Demand architecture cycles operate on shorter and longer horizons simultaneously (procurement standards can change quickly; institutional integration depth accumulates slowly). Synchronized development requires budget instruments that recognize these heterogeneous cycles and protect investments in slower-cycle layers (especially capability) from the political pressure to redirect resources toward faster-cycle visible outputs.

Third, performance metrics must reflect the multiplicative condition (the diagnostic, not estimable, structural claim of section 4.2). A state that evaluates AI strategy on compute-investment metrics alone — GPU counts, data center capacity, capital deployed — measures only one layer and rewards visibility-bias. Metrics should track inference-capacity outputs: operations executed, operational integration depth across sectors, and governance properties of the inference being conducted. The section 6 indicator system provides the metric structure; what is required is institutional commitment to use it as the primary basis for AI strategy evaluation rather than the input metrics current strategies typically employ.

9.4 Profile-Specific Prescriptions

The universal prescriptions in section 9.1 – section 9.3 apply across all non-frontier states. The framework also generates profile-specific prescriptions that vary by binding-constraint layer. We organize these by the three-layer profiles illustrated in section 7.

For Layer-I-constrained states (Estonia profile, section 7.1), the priority is governable compute access, not domestic construction. Frontier-scale data centers are structurally infeasible and unnecessary; the objective is access meeting the gross-vs-governable standard (section 2.3) — diversified suppliers, continuity guarantees, jurisdiction-protected access, and shared sovereign-compute arrangements. EU AI Factories provides one model; diversified hyperscaler arrangements under explicit governance terms; WEF's Digital Embassies for Sovereign AI concept a third. The challenge is negotiating these from positions of governance maturity rather than procurement convenience.

For Layer-II-constrained states (Azerbaijan, section 7.5), the priority is capability paced against demand architecture. Capability-export (section 5.2) is a characteristic risk. Workforce programs alone — AI academies, training, university partnerships — do not address the constraint without coupled deployment opportunities. The bundle is capability investment plus embedding mechanisms (mandatory integrator roles in ministries, MLOps career tracks in regulated sectors, public-sector deployment) plus sectoral mandates creating demand pull. Workforce and demand policy are inseparable components of a single intervention.

For Layer-III-constrained states (UAE profile, section 7.3), the priority is demanding architecture construction, with procurement reform as a primary instrument. Procurement converts demand from signal into operational market: tender specifications requiring AI-enabled delivery, performance metrics accounting for AI integration, contractual standards specifying both performance and governance properties. A state with substantial compute and growing capability that has not reformed AI procurement has built supply against demand it has not constructed; the multiplicative condition predicts diminishing returns on further compute or capability until demand catches up. Demand architecture also includes regulatory clarity on data use, sectoral mandates with performance targets, public-sector training, and metrics institutionalizing demand within routine evaluation.

For uneven profiles (Indonesia, section 7.4), the priority shifts from absolute level of any layer to consistency across the national territory. The multiplicative condition operates at the level of institutional units consuming inference — ministries, agencies, sectoral systems — and unevenness across them produces uneven returns. Instruments are sub-national capacity development, sectoral coverage beyond flagship initiatives, and institutional coverage across lagging regions. Sahabat-AI provides a structural model: locally adapted capability on domestically governable compute, scaled with attention to multilingual and sub-national contexts.

For balanced profiles (Singapore, section 7.2), the priority is interdependence governance complexity: managing multiple suppliers, jurisdictional, and sectoral integration relationships where coordination is the strategic challenge. Instruments are negotiation capacity from governance maturity, diversification against single-supplier or single-jurisdiction concentration, and analytical capacity to anticipate geopolitical fragmentation.

9.5 Sequencing

The multiplicative condition has implications for the *order* in which policy moves are made, not only their content. Two sequencing principles follow the framework.

First, diagnosis precedes investment. New strategic investments should be evaluated against the layer-profile diagnosis developed under section 9.2, not against the political salience of the proposed investment or its alignment with frontier-state strategic templates. This is the practical operationalization of the diagnostic prescription.

Second, governance specification precedes integration depth. The framework's exclusion of uncritical procurement (section 9.1) is most operationally consequential at the moment when new supplier relationships are being established. Once institutional dependence on an inference service has developed, the bargaining position of the state vis-a-vis the supplier weakens, and the governance terms of achievable post-dependence are systematically inferior to that achievable pre-dependence. The sequencing prescription is that governance properties — audit access, continuity guarantees, exit rights, jurisdictional specifications — be negotiated as primary contract terms at the inception of supplier relationships, not as retrofits after operational integration has deepened.

These principles do not specify which layer should be developed first in absolute terms (section 4.4); they specify which institutional commitments should be made first within each policy intervention: diagnose before investing; specify governance before integrating.

10. Limits, Research Agenda, and Application: The Azerbaijan Case

The framework rests on structural claims whose validity is bounded by specific conditions and whose policy implications are conditional on assumptions that bear explicit articulation. This section discusses those bounds, identifies the research program required to refine the framework further, and develops the Azerbaijan case as the worked example through which the framework's claims are most concretely applied — and through which both its diagnostic value and operational limits surface.

10.1 Boundary Conditions of the Framework

The framework's claims are bounded by three structural conditions. Each is a working assumption rather than an established fact, and each constitutes a research question whose investigation would refine the framework of policy implications.

Inference-dominant phase. The framework is specified for the period in which inference workloads dominate AI compute composition, and outputs are increasingly embedded in institutional workflows. The phase is assumed to persist through at least 2030. Technological developments restructuring composition — a return to training dominance under different architectures, or a shift toward agentic systems with qualitatively different compute profiles — would require framework revision. Policy commitments should therefore be reviewed at the cycle on which technological-phase assessments are conducted, not held permanently.

Institutional embedding. Layer III rests on the assumption that productive inference depends on institutional integration into administrative, regulatory, and organizational decision processes. If inference comes to be consumed predominantly by individual users with institutional embedding peripheral, the multiplicative condition of three-argument structure would require reconsideration. The Indonesia case (Sahabat-AI deployed across institutional and consumer use) suggests embedding remains central in current trajectories; whether this persists is open.

Cost trajectory. Per-token inference costs have declined substantially over 2023–2026. If costs decline faster than institutional integration deepens, the value of governable domestic compute relative to commodity-priced foreign inference falls; in the limit, where inference is priced as a near-zero-marginal-cost utility, the productive distinction between Layer I configurations weakens. Compute-related implications are therefore conditional on cost decline relative to integration; the posture under fast cost decline (lighter Layer I, heavier Layer III) differs from that under stable costs (balanced development), and framework-aligned states should monitor the trajectory accordingly.

10.2 The Research Agenda

The framework's contribution to policy analysis is contingent on the empirical and methodological work required to operationalize it. Four research priorities follow from the present paper.

Indicator operationalization at the country–sector level. The nine indicators in section 6 require three methodological steps: measurement protocols implementable by national statistical agencies with available data; weighting schemes that aggregate across layers consistent with the multiplicative condition without overcommitting to a functional form; and cross-country comparability standards. In-country pilots with statistical agencies in AIPI-aligned jurisdictions are the appropriate next step.

Layer interaction empirics. Whether layer interactions are best characterized as multiplicative, additive, or hybrid (with bounded but non-zero substitutability) is a substantive empirical question. Identification requires cross-country variation in layer profiles paired with measurable inference-output indicators — a data infrastructure not yet existing but implied by the operationalization work above. Stronger multiplicativity strengthens the binding-constraint logic; weaker multiplicativity admits more substitutability and weakens the universal-prescription set in section 9.

Boundary condition validation. Each boundary condition admits empirical tracking: compute composition data for the inference-dominant phase; the workflow integration index for institutional embedding; inference cost benchmarks (e.g., Epoch AI, provider disclosures) for cost trajectory. Systematic monitoring provides an early warning system for framework retirement.

Cross-paper integration. The framework operates within a broader research program — AIPI/ISI on infrastructure threshold status (Ibrahimov, 2025a), CTS on cultural prerequisites (Ibrahimov, 2025b), SINT on threshold dynamics (Ibrahimov, 2025c), the present framework on the production function, and Governed Interdependence on the institutional architecture for managing dependence (Ibrahimov, 2026). Formal integration specifying how diagnostic outputs from one inform input to another is a priority exceeding the scope of any single paper.

10.3 The Azerbaijan Case: Layer-Profile Diagnostic

Azerbaijan provides the most directly applicable test of the framework's diagnostic value: its layer profile illustrates the central diagnostic claim that the binding-constraint layer is not necessarily the most visible layer. The section 7.5 sketch positioned Azerbaijan as a transitional case with Layer II as the binding constraint; this section develops that diagnostic in greater detail, identifies sectoral variation, and surfaces the framework's operational limits in this specific context.

10.3.1 Layer I — Substantive Capacity, Strategic Underdetermination

Azerbaijan's Layer I is substantive but strategically underdetermined. The TransCaspian Fiber Optic project and Digital Silk Way position Azerbaijan as a regional digital transit hub; AzInTelecom's 2025 national HPC centre established sovereign supercomputer infrastructure; cloud partnerships with AWS, Google Cloud, and Microsoft Azure (Ibrahimov, 2026) provide diversified hyperscaler access; G-Cloud is expanding; and SABAH.city includes data center capacity in active development—all within the Concept of Digital Development (Republic of Azerbaijan, 2025b).

Together these constitute substantive Layer I, but the governance architecture converting them into a coherent compute portfolio — explicit continuity, audit, exit-rights, and jurisdictional terms — is still in development. Per the gross-vs-governable distinction (section 2.3), Azerbaijan's gross compute is substantive while governable compute is the institutionally constituted subset. The Layer I priority is therefore governance architecture, not additional capacity. The energy endowment is a strategic advantage not yet integrated with AI compute planning at the level of section 5.1's co-adequacy logic implies.

10.3.2 Layer II — The Binding Constraint

Layer II constitutes the binding constraint in the framework's most operationally consequential sense. The AI Strategy 2025–2028 (Republic of Azerbaijan, 2025a) identifies workforce development as a central priority, providing concrete instruments — the planned AI Academy, specialized training programs, dedicated Azerbaijani NLP capacity. The Center for Analysis and Coordination of the Fourth Industrial Revolution (C4IR) provides an institutional anchor, and the existing IT workforce provides a substantive base for AI-specific capability cultivation.

Three structural challenges shape the dynamics. First, the current base of AI operations professionals is small relative to the ambitions of the AI Strategy and the demand the Digital Economy Strategy implies. Second, international labor markets create persistent brain-drain risk to higher-paying EU, Gulf, and East Asian jurisdictions, making retention rate critical to net Layer II development. Third, I&I requires role-types — sectoral domain translators, ministry-embedded integrators, MLOps specialists in regulated sectors — whose cultivation requires longer embedding cycles than generic IT workforce development. SINT (Ibrahimov, 2025c) reads Azerbaijan consistently: strong policy and infrastructural pillars alongside uneven human-capital and institutional-learning layers.

The framework's policy implication, developed in section 9.4, is that operational capability development must be paced against demand architecture construction to avoid the capability-export failure mode. The AI Academy and related instruments address the capability-supply side; their effectiveness depends on the construction of domestic deployment opportunities — through procurement reform, sectoral mandates, and institutional embedding programs — sufficient to retain capability productively.

10.3.3 Layer III — Emerging Strength, Conversion Challenge

Layer III shows emerging strength alongside a conversion challenge. The ASAN service architecture (over 450 e-government services) provides an institutional substrate; the Digital Economy Development Strategy 2026–2029 (Republic of Azerbaijan, 2025c) alongside the AI Strategy 2025–2028 (Republic of Azerbaijan, 2025a) signal genuine institutional commitment; and substantial non-oil sector investment over the past two decades indicates capacity for sustained strategic implementation across multi-year cycles.

The conversion challenge is the gap, identified in section 5.3 as the *strategy-without-demand* failure mode, between high-level strategic commitment and the procurement standards, regulatory clarity, and performance metrics that constitute *constructed* rather than *signaled* demand. The SINT analysis of Azerbaijan's sectoral configuration (Ibrahimov, 2025c) provides useful disaggregation: public administration sits in a configuration of high policy intensity but moderate societal demand — the Mandate Compliance configuration — where centralized programs were institutionalized before full cross-agency data readiness; finance and energy sit in configurations of high policy intensity matched by high societal demand — Convergent Momentum — where AI integration proceeds under coordinated supply-demand alignment. The implication for the present framework is that the binding-constraint analysis at the national level partially obscures sectoral variation: in finance and energy, Layer III is converging toward productive configuration; in public administration, Layer III shows institutional commitment without operational depth.

The strategic priority at Layer III is therefore differentiated by the sector. In sectors where Layer III is converging (finance, energy), the binding constraint is Layer II capability sufficient to absorb available demand. In sectors where Layer III shows mandate-compliance configuration (public administration), the priority is the conversion of strategic commitment into procurement and regulatory architecture that creates operational rather than declared demand.

10.3.4 Asynchrony Pattern and Strategic Implications

The aggregate pattern is moderate-C, low-S, moderate-D, with significant sectoral variation. The multiplicative condition predicts Layer II as the national binding constraint, with investment in Layers I and III yielding diminishing returns until Layer II catches up. The section 9.4 prescription is an integrated capability-and-demand strategy: capability investment (AI Academy, training programs, NLP development) coordinated with demand architecture construction (procurement reform, sectoral mandates, embedding programs) so capability output finds productive deployment rather than exporting.

Governed interdependence (section 8.2) implies parallel Layer I commitments: supplier diversification, explicit governance specification in cloud arrangements, and energy-compute co-design enabled by Azerbaijan's energy endowment. Simultaneous engagement with multiple geocognitive power poles (Ibrahimov, 2026) — domestic compute, US and European cloud partnerships, Chinese-aligned digital corridor participation — is supported on the condition that governance architecture across these relationships is institutionally specified rather than left to bilateral emergence.

10.4 What the Azerbaijan Case Reveals About the Framework

Three observations about the framework itself follow from the Azerbaijan application.

First, the multiplicative condition's diagnostic value is sensitive to the level of analytical aggregation. At the national level, Layer II is the binding constraint; at the sectoral level, the binding constraint varies substantially. A diagnostic apparatus that operates only at the national level would mis-specify the policy's priority for finance and energy sectors, where Layer II is approaching adequacy, and Layer III demand absorption is the local constraint. The implication for the framework is that the indicator system in section 6 must be applied at the country–sector level, paralleling the AIPI/ISI architecture practice (Ibrahimov, 2025a), rather than at the country level alone.

Second, prescriptions are conditional on sustained implementation cycles whose stability is itself a strategic variable: the AI Strategy 2025–2028 and Digital Economy Development Strategy 2026–2029 depend on political-cycle continuity, sustained budget execution, and the institutional capacity-building SINT flagged (Ibrahimov, 2025c).

Third, governed interdependence (section 8) admits multiple operational pathways. Azerbaijan's simultaneous engagement across geocognitive power poles is consistent with the configuration; deeper alignment with a single pole could also be consistent under different governance architecture. The framework specifies the structural conditions; the choice among consistent configurations is strategic judgment.

11. Conclusion

The strategic question for non-frontier states in the AI era is not who built the largest models. It is who can sustain inference capacity, embed it into governance and economic processes, govern its integration responsibly, and maintain meaningful sovereignty within conditions of structural interdependence. This reframing — from AI capability as aggregate ambition to inference capacity as a specifiable productive output — is the analytical move that distinguishes operationally substantiated strategies from those that perform the form of frontier-state ambition without the substance.

The framework specifies these conditions through three multiplicatively interacting layers — compute and data infrastructure, operational capability, and institutional demand architecture. The binding-constraint layer varies by national context and must be diagnosed before investment is allocated; the asynchrony penalty that follows from uneven layer development is the structural explanation for the gap between AI strategy ambition and observable integration. The ISI extension (section 6) provides the diagnostic apparatus.

The shift is from AI leadership as policy objective to sustained inference access as public capability — recognition that strategic value lies in the institutional capacity to convert inference into governance quality, economic productivity, and social welfare. Governed interdependence (section 8) is the posture consistent with this shift: structured engagement under terms preserving domestic governance access, rather than autonomy maximalism (structurally infeasible) or uncritical integration (producing services without infrastructure). The framework's claims are bounded by the inference-dominant phase, institutional-embedding assumption, and cost trajectory; their monitoring is itself a strategic variable. The framework specifies the production conditions of inference; what should be done depends on national circumstances it illuminates but does not resolve.

Authors' Declaration
Author contributions. Sole author; responsible for conceptualization, analysis, and writing.

Use of AI tools. ChatGPT and Claude assisted with language editing, literature-scoping (identifying sources), and document consistency. These tools were not credited as authors and did not make independent claims. All analysis, judgments, and final text were reviewed and approved by the author, who takes full responsibility for the content.

References

1. Bresnahan, T. F. and Trajtenberg, M. (1995) 'General Purpose Technologies: Engines of Growth?', *Journal of Econometrics*, 65(1), pp. 83–108.
2. Brookings Institution (2026) *Is AI Sovereignty Possible? Balancing Autonomy and Interdependence*. Washington, DC: Brookings Institution.
3. Deloitte (2025) 'Why AI's Next Phase Will Likely Demand More Computational Power, Not Less', in *Technology, Media, and Telecommunications Predictions 2026*.
4. DiMaggio, P. J. and Powell, W. W. (1983) 'The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields', *American Sociological Review*, 48(2), pp. 147–160.
5. Federal Reserve Board (2025) 'The State of AI Competition in Advanced Economies', *FEDS Notes*. Washington, DC: Federal Reserve Board.
6. Frischmann, B. M. (2012) *Infrastructure: The Social Value of Shared Resources*. New York: Oxford University Press.
7. Garicano, L. and Rossi-Hansberg, E. (2015) 'Knowledge-Based Hierarchies: Using Organizations to Understand the Economy', *Annual Review of Economics*, 7, pp. 1–30.
8. Ibrahimov, O. (2025a) 'AI as Public Infrastructure: A Critical Review of the Transition from Tool to Societal Necessity', *Current Trends in Computing*, 3(2), pp. 40–61.
9. Ibrahimov, O. (2025b) 'Cultural–Technological Synergy in the Age of AI: A Conceptual Framework for Understanding Adaptive Modernization in Transitional Societies', *UNEC Journal of Computer Science and Digital Technologies*, 1(2), pp. 5–29.
10. Ibrahimov, O. (2025c) 'Making Intelligence Public: Thresholds of Policy, Demand, and AI-Readiness', *Journal of Modern Technology and Engineering*, 10(3), pp. 189–212.

11. Ibrahimov, O. (2026) 'Digital Sovereignty in the Emerging Global Cognitive–Informational Order: Geocognitive Power Poles and Governed Interdependence', *Journal of Modern Technology and Engineering*, 11(1), pp. 61–85.
12. Ide, E. and Talamas, E. (2025) 'Artificial Intelligence in the Knowledge Economy', *Journal of Political Economy*, 133(12).
13. Indosat Ooredoo Hutchison and GoTo (2024) *Sahabat-AI: Indonesia's Sovereign AI Initiative*. Jakarta: Indosat Ooredoo Hutchison. https://www.indosatooredoo.com/en/sahabat-ai
14. Indosat Ooredoo Hutchison and GoTo (2025) *Sahabat-AI 70B Scaling Announcement*. https://www.indosatooredoo.com/en/sahabat-ai
15. Luitse, D. (2024) 'Platform Power in AI: The Evolution of Cloud Infrastructures in the Political Economy of Artificial Intelligence', *Internet Policy Review*, 13(2).
16. McKinsey & Company (2025) *The Sovereign AI Agenda: Moving from Ambition to Reality*. New York: McKinsey & Company.
17. Norris, T., Profeta, T., Patiño-Echeverri, D. and Cowie-Haskell, A. (2025) *Rethinking Load Growth: Assessing the Potential for Integration of Large Flexible Loads in US Power Systems*. Durham, NC: Nicholas Institute for Energy, Environment & Sustainability, Duke University.
18. Republic of Azerbaijan (2025a) *Artificial Intelligence Strategy of the Republic of Azerbaijan for 2025–2028*. Presidential Decree, 19 March 2025. Baku: Office of the President.
19. Republic of Azerbaijan (2025b) *Concept of Digital Development of the Republic of Azerbaijan*. Presidential Decree, January 2025. Baku: Office of the President.
20. Republic of Azerbaijan (2025c) *Digital Economy Development Strategy 2026–2029*. Baku: Ministry of Economy.
21. Infocomm Media Development Authority (2024) *Model AI Governance Framework*. Singapore: IMDA. https://www.imda.gov.sg/
22. Star, S. L. and Ruhleder, K. (1996) 'Steps Toward an Ecology of Infrastructure: Design and Access for Large Information Spaces', *Information Systems Research*, 7(1), pp. 111–134.
23. Tony Blair Institute for Global Change (2026) *Sovereignty in the Age of AI: Strategic Choices, Structural Dependencies and the Long Game Ahead*. London.
24. World Economic Forum (2026) *How Shared Infrastructure Can Enable Sovereign AI*. Geneva.

Assessing Azerbaijan's Digital Transformation Profile Through the New Generation Economy Index 2025

Arzu Huseynova
Economic Scientific Research Institution, Ministry of Economy of the Republic of Azerbaijan, Baku, Azerbaijan
0000-0002-0981-9923, arzu.huseynova@esri.gov.az

Abstract

As national economies undergo digital transformation, there is a growing need to measure not only technological readiness but also the ability to translate digital opportunities into tangible socioeconomic changes. In this regard, this paper presents a situational analysis of Azerbaijan's position in the 2025 Next Generation Economy Index (NGEI), a multidimensional assessment framework that includes fifteen countries and ten relevant development dimensions. In line with this objective, the study uses a diagnostic model and explores the path of factors contributing to outcome achievement based on the National Digital Transformation Profile. Azerbaijan ranks eighth among 15 countries in the Next Generation Economy Index. Azerbaijan also performed well in demographic potential, education, skills, and technological readiness. However, although we note that the indicators for innovation and the creative economy were relatively low, these discrepancies indicate that the country's development potential in outcome-related areas has not been fully realized. The study's findings highlight the importance of innovation ecosystems, research commercialization, entrepreneurship, digital exports, and interinstitutional collaboration in ensuring sustainable economic transformation. This study provides a diagnostic picture of the situation in Azerbaijan and details a number of factors that could help transform existing skills into sustainable economic well-being much more quickly.

Keywords: Digital Transformation, Digital Economy, Innovation, Economic Development, Azerbaijan

Received:	*Revised:*	*Accepted:*	*Published:*
20/05/2026	*03/06/2026*	*11/06/2026*	*17/06/2026*

1. Introduction

The rapid development of digital tools has shifted the basic fundamentals of the economy. The present economy has more influence on competitiveness based on the ability to generate, transfer, and promote knowledge and not just the production and accumulation of resources [work and capital]. What we know then is that digital technologies have the potential to impact production processes, business models, public administration, innovative systems, and international trade processes [5].

The expansion of conventional indicators of economic performance presents a partial and partial picture of development as digitalization moves from one sector and enters into more widespread domains. Increased access to the internet, digital infrastructure, and technology preparedness do not automatically lead to greater creativity, entrepreneurship, and productivity. Consequently, growing attention has been focused on the extent to which countries turn digital capabilities into the tangible output of their economies [3, 6]. This represents a critical problem for developing and middle-income countries as spending on infrastructure and human capital increases faster than building innovation ecosystems and commercialization processes [9].

Some recent research notes that the technological, organizational, and institutional components were considered to be essential to successful digital transformation [3, 4]. Firm human capital development, innovation capacity and ecosystem coordination shape economy response to sustainable technological change affects the ability of economies to generate the benefits from sustainable technology change driven advantages [5, 8]. Countries with established innovation systems tend to be more productive and economic competitive, but platform-based business activities and creative fields more and more add value and often even contribute further to economic diversity at greater scales [7, 9, 10, 14].

In turn, these advances have led to the emergence of increasingly more multidimensional tools used to assess economic change. Composite indices have become increasingly popular to survey phenomena that are hard to quantify by one single index [15]. These models make it possible to study the interaction of the technologies, the economic, the social and the institutional developments, to ensure their international comparison [15].

One such index is the 2nd Generation Economy Index [1, 2]. This is particularly important in the digital transformation arena, where higher levels of preparedness are not always associated with equally positive innovation and commercialization.

In recent years, Azerbaijan has built a digital infrastructure, strengthened its educational potential, and promoted technological modernization. However, the extent to which these advances have had a positive impact on innovation, digital entrepreneurship, the development of the creative economy, and the international competitiveness of digital activities is questionable.

In this context, the aim is to determine Azerbaijan's position in the 2025 Generation Economy Index and identify the country's key strengths and weaknesses in its digital transformation, to examine Azerbaijan's situation from all perspectives at the same time, and to identify potential opportunities for the government to be more actively involved in the country's new economic transformation.

1.1 Problem statement

Today, digital transformation is a strategic priority for increasing productivity, expanding competition, and accelerating economic growth. Many resources are put into digital infrastructure, technological readiness, education, and skills development. They set the stage for participation in the digital economy. Although good conditions are available, good consequences are often not achieved [3, 4].

Progress is, in many instances, evaluated against measures related to technology and digital resources. Success is often measured by such indicators as the penetration of broadband access, the use of the Internet, digital government services, and technological infrastructure. However, one indicator is not enough to fully reflect the success of transformational processes. Countries can demonstrate a very high level of readiness, but at the same time show low results in innovation, entrepreneurship, digital export, and value creation based on knowledge [5, 8].

Studies show that economic benefits are mainly due to the ability to transform into productive activities, innovative products, and sustainable competitive advantages [3, 6]. On the other hand, the relationship between readiness and the results of the assessment of digital development is currently an urgent problem. The Azerbaijani instance is illustrative of this challenge. A lot of progress has been taken on issues like technological modernization, digital infrastructure, education, and skills. These gains have significantly enhanced the nation's participation in the new digital economy. However, existing evidence indicates that progress in innovation, creative economy activities, startup development, and technology commercialization have lagged behind [1, 2].

This situation poses an important analytical question. How are existing capabilities converted into discernable economic outcomes? The answer is significant, since there are multiple policy interventions depending on problem type or context. In the event limitations are a consequence of unpreparedness, more should focus on infrastructure, education, and adoption of technology. Where readiness already exists but consequences are poor, focus needs to be on the innovation systems, research commercialization, entrepreneurship, venture finance and ecosystem development [4, 7].

Consequently, it is the relationship between enabling conditions and achieved outcomes within the framework of the digital transformation in Azerbaijan that is addressed to be the principal problem of the current study. Understanding this relationship would clarify current strengths and weaknesses and would facilitate the prioritization of economic development issues for the future.

1.2 Literature Review

The linkage of digital technology and economic growth has emerged as one of the most studied issues in modern economic studies. The swift proliferation of digital technologies has also affected production systems, labour markets, innovation paths, corporate structures, and public services. Therefore, modern studies tend to regard a digital transformation as a multidimensional phenomenon of which it is not merely a matter of technology adoption [3].

The earlier research focused at times on information and communication technologies and productivity, and economic growth. Subsequent research moved up a ladder of broader questions about the adoption by organizations, institution building capacity, and value creation. Recent research has been showing that technology resources have meaningful impacts only were accompanied by changes in support on management, human capital, and innovation systems [4, 5].

An expanding body of literature has stressed digital innovation as one of the agents for economic transformation. Innovation is no longer considered to be the result of research-based activities. It is also being recognised as a part of the economy. It is understood instead as a process involving knowledge creation, commodity commercialisation, market reformulation, and ongoing learning from a system of knowledge processes and market evolution. The extent to which this process works also relies on the interaction among universities, the firms, the enterprises, the investors, the public institutions and other actors in a wider innovation ecosystem [6, 7].

Recent studies have also highlighted that digital platforms have begun to take a more prominent role in economic activity and the growing importance of using technology and ecosystem-based economic activity. Platform models encourage producers, consumers, and service providers to engage more with each other and also bring down transaction costs and facilitate access to markets. Their increasing influence has reconvened classical forms of competition, entrepreneurship, and international business development [14].

A second key area of research focuses on the creative economy. International reports published recently also reflect the increasingly significant influence of the ability to innovate and digital responsiveness on economic performance. As stated in the Global Innovation Index 2024, developing countries that perform well in the fusion of knowledge making, know-how and technical development and commercialization have a higher level of innovation performance and a higher international competitiveness [7]. The OECD’s Digital Economy 2024 report highlights the growing importance of digital technologies in boosting productivity, improving economic efficiency, and accelerating structural change [8].

Therefore, digital transformation should be viewed not only as a technological innovation, but also as a multidimensional development that affects the organization of economic activity, the innovation potential, and the means of value creation. The multidimensional nature of digital transformation has created the need for more complex analytical architectures to measure digital transformation. Composite indices are one of the common frameworks for assessing complex socio-economic processes that cannot be fully measured by a single indicator. These indices allow the integration of technological, economic, social, and institutional indicators into a single analytical framework and provide a more comprehensive basis for cross-country comparisons [15].

Areas such as digital infrastructure, innovation, education, entrepreneurship, and institutional quality have been extensively studied in the literature. The interaction of these areas and the translation of potential into economic impacts have been less studied. This is particularly relevant for countries where digital readiness, human capital, and technological capabilities are developing faster than innovation, commercialisation or creative economy outcomes. In this context, the Next Generation Economy Index provides an analytical framework for a broader perspective. The index allows for a distinction between enabling conditions and achieved outcomes; it also allows for the study of economic transformation based on the degree to which each resource is realised as economic value, rather than on existing resources [1, 2].

Applying the above methods to the Azerbaijani context provides a suitable theoretical context for a more systematic examination of the country's digital development potential, innovation capabilities, entrepreneurial environment and potential for sustainable economic value creation.

2. Experiments

2.1 Analytical framework and data

The research derives empirical evidence from the "Next Generation Economy" composite index. The index combines interrelated dimensions such as human capital, technological readiness, innovation potential, social development, and economic indicators into a single assessment system [1, 2].

The benefit of the composite index technique is that multiple dimensions are combined into an analysis model; however, individual analytical analysis may still be performed in terms of each one [15]. This creates a methodological base for the systematic evaluation of the digital transformation process influenced by various interacting technological, economic, social, and institutional factors.

The NGEI is made up of ten subdimensions grouped into two major categories. The first one represents the enabling conditions. Such dimensions refer to the resources and capabilities conducive to economic transformation. The second category represents realised outcomes and assesses the extent to which available capacities yield measurable economic results [1, 2].

Factors promoting participation cover demographic capacity, health, education, skills, technological readiness, social inclusion, and institutional support. Taken together, the indicators indicate the structural aspects essential for participation in the digital economy. Outcome dimensions encompass innovation performance, creative economy development, and platform economy activity. These measures demonstrate the financial benefit that results from the efficient utilization of resources and skills in a location [1].

The framework relies on the assumption that strong outcomes will be more likely to take place when enabling conditions are adequately developed. Simultaneously, favorable conditions are not inherently good for success. Variations between these two pairings of indicators may thereby identify obstacles in the transformation process and useful information for policy discussion [3, 4].

The current research uses diagnostics instead of causal analysis. It does not seek to estimate econometric relationships or identify direct causal effects between independent variables. Rather, the analysis considers the internal structure of Azerbaijan's NGEI profile, and it compares the readiness-related factors with the outcome-related dimensions. We use this approach to pinpoint areas where development seems to be relatively strong and areas that need further work. The empirical analyses depend on the data reported in the official New Generation Economy Index 2025 Global Ranking Report as well as the methodological documentation [1, 2].

These sources offer information on performance, subdimension scores for total performance, country rankings, and distribution of results by input and output category. This article is based on data reported which are the parameters for the measurement. The structure of the framework in focus on the relationship between enabling conditions and realized outcomes provides a more holistic approach to the economics of transformation than do traditional ranking techniques. It gives you the opportunity to pay attention not just to the overall performance, but to the process of how development capabilities are turned into economic value that is quantifiable.

3. Results and discussion

3.1 Azerbaijan's diagnostic profile

Azerbaijan achieved 42.11 points on the New Generation Economy Index 2025 and placed eighth out of the fifteen countries to take part in the evaluation [1]. Despite the relatively general status of results, further analysis of the index's internal dimension shows some crucial elements in the profile of national development of the country. Demographic capacity, skills, technological readiness, and education show the strongest results, as highlighted in Table 1.

These features of the digital economy may suggest a broadly favorable environment to engage in the digital economy. The results of the report in this section imply that a majority of the essential criteria concerning technological and economic adaptation in Azerbaijan have been adequately established [1]. When the outcome-related dimensions are considered, however, an alternative picture is revealed. Innovation in the national profile, as indicated in Table 1, ranks lowest, and the creative economy shows a fairly moderate performance. Platform-economy activity outperforms those benchmarks but has not reached the thresholds set by the country's readiness indices [1].

The comparison of preparedness to outcomes presented in Table 1 is one of the key issues in this work. Capacity seems to exceed the economic value of those capacities. This mismatch tells us that there is not so much existing talent in place for innovation, entrepreneurship, knowledge-based value creation, and this value is not being maximized. The findings suggest that Azerbaijan's main limitation is not due to lack of technology access or human capital development. The issue lies in the degree to which resources, including human capital, knowledge, and technological capabilities, are translated into economic returns. In this respect, the index structure presented in Table 1 is not just consistent with the country's position in the world, but it also represents the institutional and functional deficiencies of turning digital transformation potential into economic returns.

Table 1. *Azerbaijan's NGEI diagnostic profile*

Indicator	**Value**
Overall NGEI score	42.11
Rank	8/15
Input score	34.12
Output score	8.00
Profile	Medium-high input / low output
Demographic score	82.7
Skills and technological readiness	63.9
Education score	53.6
Innovation score	6.5
Creative economy score	17.1
Platform and sharing economy score	34.6

Source: Compiled by the author based on NGEI 2025 Global Ranking Report [1].

3.2 Interpreting the observed pattern

The above findings also indicate a wide-spread gap between the enabling conditions for innovation, as well as its performance and commercialization results. This divide exists across many developing countries and middle-income economies which invest more in education, digital infrastructure, and technology readiness than in the outcome and commercialization of innovation [3, 4]. Another explanation is that countries may be competent for knowledge generation but poorly developed for knowledge commercialization, even when they possess the appropriate human capital and institutional systems to create enough opportunities for innovation. A good link between research organizations, universities, businesses, investors, and other institutions is also required for the knowledge to be commercially valuable [6]. A further key influence is entrepreneurship. In business, start-ups create new technology or new products at a higher rate and develop markets. So, value from the start is not limited to their own economic activity. They usually provide more competition, investment from the top down, and the dissemination of technology, which in principle buffers the effects of technology and education in a thin startup ecosystem [7].

The interpretation of these results rests on the notion of digital value creation. We know that the economic impact of technology comes from its embedding in production, or indeed in the business process itself [4, 5]. According to the data, the process of digital transformation is not only about the adoption of new technology but also about the establishment of sustainable economic activities that raise productivity, increase innovation, and offer a competitive edge. Also, the low performance of capabilities per innovative and creative economy dimensions means that their economic benefits are too low. In order to promote innovation, the state of readiness cannot be expected to facilitate this [7, 8, 9]. The overall conclusion of the assessment is that while Azerbaijan is making progress to lay the ground of the future for digital development, it is nevertheless challenged to create value relevant to this space and to be able to monetise these things. These findings imply that infrastructure-related or access indicators are not the only global enablers of digital transformation; it's also important to be able to transform capabilities into concrete economic benefits. The preparedness and outcomes distinction thus offers a good framework for analyzing the longer-term development of production systems and their innovation and economic impact.

3.3 Policy implications

According to the analysis, there are points that lead to a more balanced development path. The key policy targets are shown in particular from Table 2, with the gaps in supporting environment and performance most evident with regard to where they are concentrated.

Table 2. *Priority bottlenecks and policy levers for Azerbaijan*

Bottleneck	Evidence from the report	Policy lever
Weak innovation commercialization	Innovation score 6.5; R&D expenditure 0.18% of GDP	Tax incentives, venture co-investment, stronger research-to-market pipelines
Thin startup base	Around 100 active startups	Seed and growth finance, incubators, procurement sandboxes
Low digital export depth	ICT exports below 1% of non-oil exports	Export support, e-commerce logistics, payments and certification infrastructure
Weak creative-economy monetization	Creative economy scores 17.1	Accelerators, licensing support, global market access for digital content
Moderate platform development	Platform and sharing economy score 34.6	Interoperable digital markets, SME onboarding, regulatory clarity

The first is to enhance the capacity for innovation. An increased focus on research commercialization has the potential to enhance the translation of scientific knowledge for economic worth. Increased cooperation between universities, research organizations, and industry could assist in pursuing this goal by connecting knowledge generation to market deployment more closely if the sector can contribute to achieving the development of a stronger link between knowledge generation and market application [6]. Based on the policy solutions introduced on Table 2, the evidence for how policy instruments related to innovation financing, commercialization as well as research-market linkages could potentially serve as one means of addressing this issue could be proposed.

The second priority is entrepreneurship and business development. New firms are frequently significant sources of innovation and economic dynamism. Economic initiatives to assist in the startups, startup acceleration, mentorship, and financial access could boost the degree to which the economy translates technological potential into financial success. In order to contribute to the building of a more active innovation ecosystem, a combination of further proliferation of seed finance, incubation models, and business support systems will be implemented (see Table 2). Further development of mechanisms of financing is an additional focal area. Innovation projects are often uncertain and have a long investment horizon. Consequently, venture capital, seed investment, and other sources of risk finance are frequently linked to superior innovation performance. Further implementation of those mechanisms could enhance participation and experimentation with new business models for tech firms [4].

Digital exports need more attention. International experience has shown that digital exports have a crucial role in contributing to innovation, technology upgrading, and productivity enhancement. Access to global digital markets leads to learning, scale-up, and transfer of knowledge in ways that cannot happen in domestic markets. Therefore, greater participation in the export sector might back up the construction of more competitive and higher value economic activities [9, 11]. The policy measures included in Table 2, also further underline the necessity of export support instruments and digital market integration.

Further benefits might come from better coordination between public institutions, educational institutions, businesses, investors, and technology intermediaries. Digital transformation is a complex and varied effort, and the activities of multiple stakeholders affect the effectiveness of development policies. Alignment of these actors may help with implementation of policy and could reinforce the innovation ecosystem as a whole [3, 5].

Lastly, monitoring systems should include outcome driven rather than readiness tracking in the focus of the monitoring system. There still need to be infrastructure, education and technology capacity development improvements. But how well these investments work should also be considered in the context of innovation, entrepreneurship, productivity and value creation. This would have a broader foundation for evidence-based policy making and for measuring developmental evolution over the long-term.

Combined, the policy directions outlined in Table 2 could help to reduce the disparity between potential capabilities and actual results. A stronger integration of these dimensions would make existing resources more effective and can further contribute to the efficiency and long-term competitiveness of the Azerbaijani economy.

4. Conclusion

This paper examined the position of Azerbaijan in the New Generation Economy Index 2025 and examined the correlation between the enabling conditions and outcomes. Based on the conclusions of these investigations, a clear base for participation in the digital economy is found; this is indeed the case in the country. There are good impacts on demographic ability, higher education, skills, and technology readiness, indicating development resources played a significant role [1].

On the other hand, its conclusions also indicate that outcome-related dimensions are not well developed. In supporting dimensions: innovation, the performance of the creative economy and platform activity do not seem to represent sufficiently the readiness of at least the most critical level yet. The dichotomy between these types of indicators is also the main result of the study and informs what phase of economic transformation in Azerbaijan we have managed to arrive at, respectively. Our results would suggest that progress in future will depend not only on the continued investment of infrastructure and human capital, but also on the effectiveness of structural devices to transform raw inputs into wealth and economic value. It will lead to the realization of this goal through a more robust innovation environment [4, 6, 7], as well as enhancing commercialization mechanisms, expanding entrepreneurial efforts, and widening international reach.

One of the key contributions to the research is the emphasis on the association between readiness and results. Although, as our survey demonstrates, such traditional evaluations will tend to emphasize aggregated performance, the analytical model will offer insights into internal differences buried in the summaries.

This view presents a holistic picture of specific strengths, weaknesses, and objectives. What we find is Azerbaijan has indeed many of the elements required to participate in the digital economy in the right way. But the wealth that benefits from such conditions remains below their full potential. Future outcomes will depend on the success of innovation instruments, commercialization strategies, entrepreneurial activity, and connections within the international market.

This article examines only the New Generation Economy Index 2025 indicators which are used to analyze the data and the framework applied for analyzing the data. Subsequent studies can further analyze the conditions that drive transformation outcomes via sectoral, comparative, and longitudinal approaches. Research of this sort would lead to more insight into the issues influencing economic change in a digital era.

Acknowledgment

No external funding was received for this research. The author acknowledges the Economic Scientific Research Institute as a contributor to the construction of the New Generation Economy Index framework and access to the analytical materials utilized in this work.

Authors' Declaration

Conflict of Interest. The author declares that there is no conflict of interest related to this study.

Authors' Contribution Statement

Arzu Huseynova contributed to the conceptualization of the study, development of the analytical framework, data interpretation, manuscript preparation, revision, and final approval of the article.

References

1. Huseynova, A. D. (2025) New Generation Economy Index 2025: Global Ranking Report. Baku: Economic Scientific Research Institute.
2. Huseynova, A. D. (2025) New Generation Economy Index 2025: Methodology Report. Baku: Economic Scientific Research Institute.
3. Vial, G. (2019) Understanding digital transformation: a review and a research agenda. Journal of Strategic Information Systems 28(2), 118, 144.
4. Verhoef, P. C., Broekhuizen, T., Bart, Y., Bhattacharya, A., Dong, J. Q., Fabian, N., & Haenlein, M. (2021) Digital transformation: A multidisciplinary reflection and research agenda. Journal of Business Research, 122, 889-901.

5. Plekhanov, D., Franke, H., & Netland, T. H. (2023) Digital transformation: A review and research agenda. Comparative strategy sensemaking and the impact of disaster on triage. European Management Journal, 41(6), 821-844. doi:10.1016/j.emj.2022.09.007
6. Nambisan, S., Lyytinen, K., Majchrzak, A., & Song, M. (2017) . Digital innovation management: Reinventing innovation management research in a digital world. MIS Quarterly 41 (1), 223-238.
7. World Intellectual Property Organization. (2024) . Global Innovation Index 2024: Unlocking the Promise of Social Entrepreneurship. Geneva: WIPO. doi:10.34667/tind.50062
8. OECD. (2024) . OECD Digital Economy Outlook 2024. Paris: OECD Publishing.
9. UNCTAD. (2024) . Digital Economy Report 2024: Shaping an Environmentally Sustainable and Inclusive Digital Future. Geneva: United Nations.
10. UNCTAD. (2024) . Creative Economy Outlook 2024. Geneva: United Nations.
11. International Telecommunication Union. (2024) . Facts and Figures 2024: Measuring Digital Development. Geneva: ITU.
12. European Commission. (2024) . Digital Economy and Society Index 2024 Report. Brussels: European Commission.
13. Acemoglu, D., and Johnson, S. (2023) . Power and Progress: Our Thousand-Year Struggle over Technology and Prosperity. New York: PublicAffairs.
14. Nambisan, S., Zahra, S. A., & Luo, Y. (2024) . Global platforms and ecosystems: Implications for international business theories. Journal of International Business Studies 55(1): 1-21.
15. OECD, European Commission & Joint Research Centre. Handbook on Constructing Composite Indicators: Methodology and User Guide. 2008. Paris: OECD Publishing. https://doi.org/10.1787/9789264043466-en

The AI Divide: How Automation Deepens Digital Inequality

Aflatun Mammadov
Department of Digital Technologies and Applied Informatics
Azerbaijan State University of Economics (UNEC), Baku, Azerbaijan
0009-0000-5355-0493, aflatun_mammadov@unec.edu.az

Abstract

AI has worked its way into the machinery of everyday decisions, who is granted credit, which diagnosis a patient is given, whether a public service can be reached at all. Its benefits are not shared evenly. If anything, the distance between the countries that build these systems and those that only use them keeps widening rather than closing. This paper treats that distance as the AI Divide and reads it as a structural condition, not a passing lag. Five mechanisms feed it, and each one feeds the next: uneven access to infrastructure; labour disruption that lands hardest on developing economies; algorithmic bias that carries old discrimination into new tools; computing power gathered into a few hands; and an environmental cost paid largely by those who caused least of it. The evidence comes from a structured review of peer-reviewed work between 2016 and 2026, set against data from international institutions, and the numbers are not subtle. Adoption in richer economies runs about three-quarters higher than in poorer ones. Some 92 million jobs are expected to go by 2030, most of them where the means to retrain people are thinnest. Facial-recognition error climbs toward 35 percent for darker-skinned women while staying under 1 percent for lighter-skinned men. Three firms hold close to two-thirds of the world's clouds. Because no strand moves on its own, the divide tends to widen itself unless policy intervenes. The paper closes with five governance principles, universal access, data sovereignty, representative development, algorithmic accountability, and international cooperation, and asks what each would mean for Azerbaijan and the South Caucasus, where the problem is already here rather than hypothetical.

Keywords: Artificial Intelligence, Digital Inequality, Algorithmic Bias, AI Governance, South Caucasus

Received:	*Revised:*	*Accepted:*	*Published:*
20/05/2026	*03/06/2026*	*06/06/2026*	*17/06/2026*

1. Introduction

Artificial intelligence (AI) has left the laboratory. Only a few years ago it was a research subject; now it helps settle who gets a mortgage, tells a court which defendants to treat as high-risk, and steers where policing falls heaviest. Its commercial footprint has grown to match. Valued at roughly 538 billion US dollars in 2023, the global AI market is expanding at a compound annual growth rate of about 19 percent [1]. In the United States, private AI investment reached 67.2 billion US dollars that same year, around 8.7 times China's figure, and by 2024 the tools had passed a billion users [2]. Most of those users, though, are in developing countries that do not yet have the governance, infrastructure, or institutions to handle the risks that travel with them [3].

What makes AI different from earlier technological waves is a feedback loop baked into the way it works. Data sharpens the models; sharper models pull in more users, and those users' hand back still more data, which sets the cycle off again. Electricity and the internet spread unevenly too, but in the end, they reached most of the world. AI may not, and the reason is that the loop pays most for whoever is already ahead. A country with less data trains weaker tools, gets less back from them, and slips a little further behind on each turn of the wheel. Kanni Wignaraja of the United Nations put it plainly in late 2025: AI is racing ahead while many countries are still at the starting line [3].

For Azerbaijan the question is practical, not theoretical. The country has already built real digital infrastructure, through the ASAN (Azerbaijan Service and Assessment Network) platform and a string of e-government strategies, and that work settled a basic point: deliberate public investment in services pays off. It now faces a harder choice, namely, how to fold AI into those services before the rules governing them exist. So, the AI Divide is a governance problem in the present tense, not a forecast. The rest of the paper maps its five connected dimensions and locates Azerbaijan's exposure on each; from there it builds five governance principles on the same evidence.

2. Methodology and Analytical Framework

For a long time, work on digital inequality fixed on connectivity, on who can get online and who cannot [4]. With AI that framing is too narrow. The divide cuts through several layers of the technical and social system at once, and five mechanisms drive it. Access to infrastructure is uneven to begin with. That first gap shapes the next: labour disruption gathers in the Global South, while algorithmic bias works its way into decisions that carry real stakes. A fourth mechanism is concentration, since AI capability and the data feeding it have pooled in a handful of countries. The last is environmental, a burden shared unevenly. Table 1 lays out this framework and, for each mechanism, sets down where Azerbaijan stands.

Table 1. *Five-dimensional AI Divide framework with Azerbaijan exposure assessment.*

Dimension	Core mechanism	Global evidence	Azerbaijan exposure
Access	Infrastructure gap; 10.6 pp North-South adoption differential (H2 2025)	24.7% North vs. 14.1% South; 2.6 billion people offline; Africa holds 1% of data-centre capacity	About 18% adoption (middle tier); Azerbaijani underrepresented in AI training corpora
Labour	92 million jobs displaced by 2030 (WEF); up to 40% of jobs exposed (IMF)	170 million new roles, mainly in the Global North; structural geographic mismatch	Outsourcing, administration, and manufacturing at high risk; limited reskilling provision
Bias	Automated systems reproduce and amplify structural inequalities	About 35:1 facial-recognition error disparity; roughly double COMPAS false-positive rate for Black defendants	No AI audit mechanism; governance gap in ASAN public services
Geopolitics	Three US firms control about 65% of global cloud; data colonialism dynamics	US private AI investment 8.7 times China's; concentration of foundational models	Near-total reliance on foreign AI platforms; limited regulatory capacity
Environment	Data-centre demand 415 TWh (2024) rising	AI carbon footprint of 32.6 to 79.7 Mt CO2	Climate-exposed South Caucasus; contributes

	toward 945 TWh (2030); asymmetric cost burden	by 2025; costs fall on low-capacity regions	negligibly yet bears disproportionate costs

Source: author's framework, drawing on [2], [3], [5], [6], [7]. CO_2 = carbon dioxide; COMPAS = Correctional Offender Management Profiling for Alternative Sanctions; pp = percentage points.

The review followed PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) principles, short of a full meta-analysis. Searches ran across Scopus and Web of Science for peer-reviewed work from 2016 to 2026, combining terms like "AI inequality," "digital divide," "algorithmic bias," "AI governance," and "automation displacement" with "developing countries" or "Global South." A study earned its place only when it was empirical or policy-analytical, appeared in English in a peer-reviewed venue, and bore on at least one of the five dimensions; opinion pieces, editorials, and high-income-only work with no comparative angle dropped out. The opening pass returned 412 records. After duplicates were removed and the rest screened, 136 remained; these shaped the synthesis below, with the most representative and recent among them cited directly. Those studies were read alongside institutional datasets, from the United Nations Development Programme (UNDP), the World Economic Forum (WEF), the International Energy Agency (IEA), the International Labor Organization (ILO) and the International Monetary Fund (IMF) to Stanford University's Human-Centered AI institute (HAI) and Oxford Insights. None of the five mechanisms, in the end, sits in a box of its own. Being shut out of access leaves a workforce more exposed to automation, and the same concentration of computing power that hardens algorithmic bias also pushes environmental costs onto the communities least able to carry them. Figure 1 follows that chain.

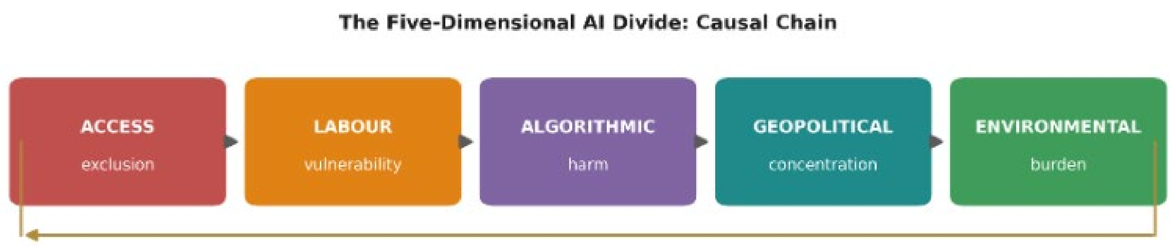

Figure 1. The five-dimensional AI Divide as a causal chain. Each dimension compounds the next, and the loop reinforces itself in the absence of deliberate policy intervention. Source: author.

3. Results and Discussion

3.1 Access and infrastructure

By the second half of 2025, the split was plain. Working-age adoption averaged 24.7 percent across the Global North and 14.1 percent across the Global South, a gap of 10.6 percentage points, up from 9.8 only six months earlier [8]. The infrastructure behind that gap is just as lopsided. Africa holds about 1 percent of global data-centre capacity, and cloud adoption there sits near 15 percent against 72 percent in Europe [9]. Table 2 gives country-level figures, and Figure 2 shows the distribution.

Table 2. *Global AI adoption rates by country, H1 to H2 2025 (%). Azerbaijan (estimate) is shown as a structural middle tier.*

Country / region	H1 2025 (%)	H2 2025 (%)	Status
UAE	n/a	64.0	Global North leader
Singapore	n/a	60.9	Global North leader

Norway	n/a	45.3	Global North
Global North average	22.9	24.7	Widening gap
Azerbaijan (estimate)	~16	~18	Middle tier
Global South average	13.1	14.1	Widening gap
Pakistan	n/a	9.6	Global South
Ethiopia	n/a	7.3	Global South, low
Malawi	n/a	4.2	Global South, low

Source: Microsoft AI Economy Institute [8]; Oxford Insights [10]; Stanford HAI [2].

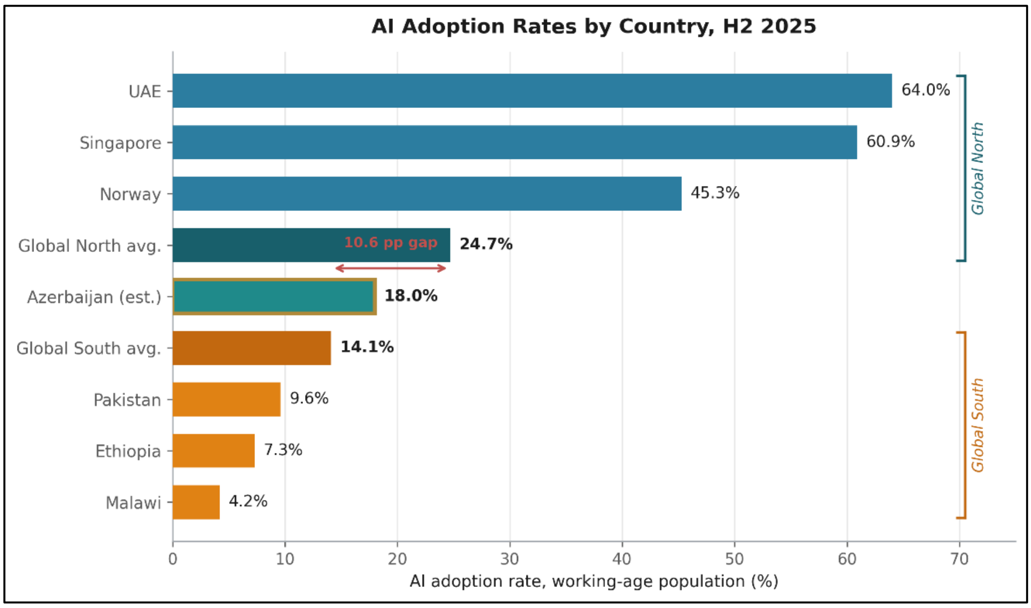

Figure 2. AI adoption rates by country, H2 2025. Azerbaijan (estimate) sits in the structural middle tier; the 10.6 pp North-South gap widened from 9.8 pp in six months, which indicates an accelerating rather than static divide. Source: [8], [10].

That leaves Azerbaijan in an uncomfortable middle: ahead of most of sub-Saharan Africa, far behind the front-runners, and not standing still. One handicap is worth pulling out. Azerbaijani barely appears in the text corpora that train large models, so tools built mainly on other languages simply work less well for the people who speak it. Worse, each new system trained on the same thin data carries the penalty forward [11].

3.2 Labour-market disruption

The WEF projects 92 million jobs displaced by 2030, set against 170 million created [5]. The IMF, looking at the same horizon, reckons generative AI could touch 40 percent of jobs worldwide [12]. The net figure flatters the picture. The roles that vanish are mostly in developing economies. The ones that appear in cluster where the infrastructure to absorb them is already in place [13]. Table 3 sets out displacement risk by sector with its relevance to the South Caucasus.

Table 3. *Sector-level AI displacement risk with South Caucasus relevance.*

Sector	Risk level	AI mechanism	Azerbaijan / South Caucasus relevance
Data entry / clerical	85%, very high	Document processing; RPA	Large clerical workforce; limited digital upskilling infrastructure

Customer service / BPO	78%, high	Conversational AI; automated resolution	Outsourcing sector growing; a major employment pathway at risk
Light manufacturing	62%, high	Robotics; automated quality control	Key export sector for transition economies; strategic vulnerability
Legal / paralegal	45%, moderate	Large language models for document review	Growing professional-services sector; risk emerging
Healthcare support	28%, lower	Diagnostic AI augments; human oversight retained	Augmentation more likely than acute structural disruption

Source: McKinsey Global Institute [14]; ILO [15]; WEF [5]. BPO = business process outsourcing; RPA = robotic process automation.

Azerbaijan's exposure is concentrated, and the list is short: business-process outsourcing (BPO), routine administration, labour-intensive manufacturing. All three are sectors with thin reskilling provision. The scale of the response required is sobering. Research from the OECD puts the cost of retraining workers in high-automation-risk jobs somewhere between 1 and 4.5 percent of gross domestic product (GDP). Actual public spending on training, averaged across OECD countries, sits at about 0.1 percent [16]. The distance between those two figures is the whole problem. The Digital Trade Hub initiative is a useful start, but at the current pace it will not build reskilling capacity before the displacement arrives. Comparable economies make the stakes clear: the ILO estimates that 26 to 38 percent of jobs in Latin America and the Caribbean are exposed to generative AI [15], and the South Caucasus profile is structurally similar. Working methodologies for assessing such risks at the level of individual firms are beginning to appear in the regional literature [17].

3.3 Algorithmic bias and discrimination

In the Gender Shades study, facial-recognition error rates climbed to roughly 35 percent for darker-skinned women while staying below 1 percent for lighter-skinned men, a gap of about 35 to 1 [18]. The Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) recidivism tool, used in several US jurisdictions, flagged Black defendants with false positives at around twice the rate of white defendants [19]. A widely used healthcare algorithm turned out to understate the needs of Black patients in a consistent, patterned way [20]. None of this comes down to a stray bug in the code. It is what tends to happen when systems are built by similar-looking teams on training data that does not represent the people the system will end up judging, a pattern long documented in the fairness literature [21], [22] and in the technical bias surveys that followed [23], [24], [25].

Azerbaijan currently has no mechanism for auditing AI systems. As automated decisions enter public services through ASAN, a citizen harmed by a biased decision has no clear route to challenge it [26]. The European Union (EU) AI Act, adopted in 2024, offers a ready template: conformity assessments for high-risk systems, independent audit rights, and a genuine right of redress [27]. Azerbaijan need not design a framework from nothing; it needs the political will to adopt one.

3.4 Geopolitical concentration and environmental burden

Three US corporations, Amazon, Microsoft, and Google, run about 65 percent of the world's cloud infrastructure [28]. For a smaller state that is a sovereignty problem as much as an economic one: the systems that increasingly shape government decisions, credit scoring, and hiring belong to firms that answer no authority in Baku [29]. The environmental side is just as lopsided. The world's data centres drew roughly 415 terawatt-hours (TWh) of electricity in 2024, and the IEA expects that figure to climb toward 945 TWh by 2030, close to Japan's entire current demand [6]. By 2025, the carbon attached to all of this may sit somewhere between 32.6 and 79.7 million tonnes [7]. The South Caucasus, meanwhile, is one of the regions most exposed to drought, heat, and food insecurity, even though it adds almost nothing to those emissions. Table 4 maps the resulting governance gap across the five dimensions.

Table 4. *AI governance gap: Azerbaijan against global benchmarks across all five dimensions.*

Dimension	Global benchmark	Azerbaijan current status	Priority action
Access	Universal broadband and AI-infrastructure mandates	About 18% adoption; ASAN operational	Extend ASAN to AI-powered public services (P1)
Labour	Reskilling investment scaled to automation risk (OECD)	Digital Trade Hub underway; scale insufficient	Accelerate outsourcing and manufacturing reskilling (P3)
Bias	Mandatory algorithmic impact assessments (EU AI Act)	No AI audit; no bias testing in public services	Enact an algorithmic accountability law; audit ASAN (P4)
Geopolitics	Data localisation and digital-sovereignty legislation	Near-total foreign-platform dependency; no data law	Pass data-sovereignty legislation; fund Azerbaijani AI data (P2-P3)
Environment	Binding energy-efficiency standards (IEA)	No AI environmental-reporting requirement	Adopt IEA standards; build climate-finance mechanisms (P5)

Source: author's analysis, drawing on the EU AI Act (2024) [27]; Global Digital Compact (2024) [30]; IEA [6]; OECD [16].

3.5 Five principles for equitable AI governance

A problem of this scale calls for governance designed to match it. The five principles below are not aspirations; each answers a documented gap and points to a concrete next step. The question of how AI comes to count as public infrastructure, and which cultural and institutional conditions allow that to happen in transitional economies, has itself become a subject of regional scholarship [31]. Table 5 summarises the principles with their application to Azerbaijan.

Table 5. *Five principles for equitable AI governance with Azerbaijan-specific applications.*

Principle	Core content	Azerbaijan application
P1. Universal access	AI infrastructure as a public good; universal-service mandates; state investment in underserved communities	ASAN is a proven precedent that targets public investment works. Extend the model to AI-powered public services at scale.
P2. Data sovereignty	Community rights over locally generated data; required representation of minority languages in training datasets	Addresses the underrepresentation of Azerbaijani in AI corpora, a gap with concrete performance consequences. National data sovereignty legislation is needed.
P3. Representative development	Public investment in local AI research capacity; linguistically diverse training data as a public good	Fund Azerbaijani, Georgian, and Armenian training data as a regional public good; build indigenous research capacity.
P4. Algorithmic accountability	Mandatory impact assessments for high-risk AI; independent audit rights; a genuine right of redress	Critical as AI enters ASAN services. Adopt the EU AI Act framework before deployment expands further.
P5. International cooperation	Binding multilateral frameworks; equitable technology transfer; redress for asymmetric environmental costs	Active engagement in the Global Digital Compact and EU AI Act processes would strengthen governance capacity and amplify Azerbaijan's voice.

Source: author's framework, drawing on [3], [6], [27], [30].

3.6 Limitations

A few limitations should be kept in view. The Azerbaijan adoption numbers in Table 2 and Figure 2 are estimates, inferred from the country's regional position rather than measured in a dedicated national survey, so they are best read as indicative. Several of the headline indicators, too, come from institutional reports and industry trackers rather than peer-reviewed studies; this grey literature is current but not refereed, and it was used only where no comparable academic dataset exists, with the source named each time. Finally, this is a single-author synthesis of a field that keeps moving. The screening followed PRISMA principles but does not extend to a full meta-analysis with a stage-by-stage exclusion diagram, and the newest figures will keep shifting. None of these changes the direction the evidence points in, but each marks a spot where a reader should stay cautious and where later work could add precision.

4. Conclusion

Read together, the five dimensions all point one way. Adoption in the Global North is running about three-quarters ahead of the Global South, and the lead is stretching [8]. Around 92 million jobs are set to go by 2030, mostly in lower-income countries, while the bulk of the 170 million new one lands where the infrastructure already is [5]. The gap in facial-recognition accuracy is no glitch; it follows how the systems were put together [18]. Most of the world's cloud capacity sits with a handful of firms, and smaller states lean on them by default [28]. And the environmental bill falls hardest on the people who get the least out of the technology [6].

For Azerbaijan, none of these stays abstract. It shows up in tools that handle Azerbaijani badly, in outsourcing and clerical staff who face automation with nowhere obvious to retrain, in public services that already lean on automated decisions while no one can be held to account for them, and in a near-total dependence on foreign AI platforms. As the UNDP warned in late 2025, without deliberate choices AI could undo half a century of development gains [3].

The picture is not uniformly bleak. Mongolia, for one, has launched a National AI Programme that is training thousands of teachers and specialists across all twenty-one of its provinces, with the stated aim of widening access and closing the gap between town and countryside [32]. Its home-grown Mongolian-language model, Egune, is a reminder that a smaller country can build for a low-resource language directly instead of waiting on foreign systems [33], which is exactly the step Azerbaijan will need to take for Azerbaijani. Bhutan has gone a similar route, standing up a national AI laboratory and development centre to grow capability at home rather than buy it in wholesale [34]. And Azerbaijan has already shown, with ASAN, that patient public investment in digital services can reach people fairly. The thread running through these cases is plain enough: inclusive AI is not really waiting on a technical breakthrough. It is waiting on choices.

Those choices are what the paper is finally about. The AI Divide was built out of decisions, about who funds the technology, whose data trains it, who writes its rules, and who is in the room when those rules are written, and decisions of that kind can be taken differently. The five principles offered here are a place to start, not a finished blueprint. The room to act on them is open today; whether it stays open is not guaranteed.

Acknowledgment

The author thanks the United Nations Development Programme, the International Energy Agency, the World Economic Forum, the International Labour Organization, Stanford University's Human-Centered AI institute, Oxford Insights, and the Microsoft AI Economy Institute for the open-access data and reports drawn on in this study. A condensed version of this work was presented at the IV International Conference on Modern Trends in Digital Economy (DEMCRO 2026) at UNEC, Baku, in April 2026; that conference paper and the present manuscript are related but separate contributions, and this manuscript is the full research submission. Generative AI tools were used for language editing and formatting; the author reviewed and verified all content and took full responsibility for it.

Authors' Declaration

Conflicts of Interest: The author declares no conflict of interest. Funding: No external funding was received for this study.

References

1. Precedence Research: Artificial Intelligence Market Size, Share and Trends 2024 to 2034. Precedence Research (2024). https://www.precedenceresearch.com/artificial-intelligence-market
2. Stanford Institute for Human-Centered AI: Artificial Intelligence Index Report 2025. Stanford University (2025). https://aiindex.stanford.edu
3. United Nations Development Programme: The Next Great Divergence: Why AI May Widen Inequality Between Countries. UNDP, December 2025. https://www.undp.org/publications/next-great-divergence
4. Van Dijk, J.A.G.M.: The Digital Divide. Polity Press, Cambridge (2020). ISBN 978-1-509-53446-1
5. World Economic Forum: The Future of Jobs Report 2025. WEF, Geneva (2025). https://www.weforum.org/publications/the-future-of-jobs-report-2025/
6. International Energy Agency: Energy and AI. IEA Publications, Paris (2025). https://www.iea.org/reports/energy-and-ai

7. de Vries, A.: The Carbon and Water Footprints of Data Centres and AI. Patterns 6(5), 101149 (2025). https://doi.org/10.1016/j.patter.2025.101149
8. Microsoft AI Economy Institute: Global AI Adoption in 2025: H2 Report. Microsoft Corporation (2026). https://www.microsoft.com/en-us/worklab/ai-economy-institute
9. International Telecommunication Union: Facts and Figures 2024. ITU Publications, Geneva (2024). https://www.itu.int/hub/publication/d-ind-ict_mdd-2024/
10. Oxford Insights: Government AI Readiness Index 2025, 8th edn. Oxford Insights (2025). https://oxfordinsights.com/ai-readiness/ai-readiness-index/
11. Bender, E.M., Gebru, T., McMillan-Major, A., Mitchell, M.: On the Dangers of Stochastic Parrots. In: FAccT '21, pp. 610-623 (2021). https://doi.org/10.1145/3442188.3445922
12. International Monetary Fund: Gen-AI: Artificial Intelligence and the Future of Work. IMF Staff Discussion Note SDN/2024/001 (2024). https://doi.org/10.5089/9798400262487.006
13. Acemoglu, D., Restrepo, P.: Tasks, Automation, and the Rise in U.S. Wage Inequality. Econometrica 90(5), 1973-2016 (2022). https://doi.org/10.3982/ECTA19815
14. McKinsey Global Institute: The Economic Potential of Generative AI. McKinsey & Company (2023). https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai
15. International Labour Organization: Generative AI and Jobs: A Global Analysis. ILO Working Paper No. 96 (2024). https://www.ilo.org/publications/generative-ai-and-jobs-global-analysis
16. Organisation for Economic Co-operation and Development: OECD Skills Outlook 2025 and the OECD AI Policy Observatory. OECD Publishing, Paris (2025). https://www.oecd.org/skills/
17. Hajizade, S., Balayev, R.: Working Methodology for Assessing Digital Transformation Risks. Journal of Computer Science and Digital Technologies 1(1), 83-89 (2025). https://doi.org/10.30546/UNECCSDT.2025.01.083
18. Buolamwini, J., Gebru, T.: Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. Proceedings of Machine Learning Research 81, 1-15 (2018). https://proceedings.mlr.press/v81/buolamwini18a.html
19. Angwin, J., Larson, J., Mattu, S., Kirchner, L.: Machine Bias. ProPublica (2016). https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing
20. Obermeyer, Z., Powers, B., Vogeli, C., Mullainathan, S.: Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations. Science 366(6464), 447-453 (2019). https://doi.org/10.1126/science.aax2342
21. Noble, S.U.: Algorithms of Oppression: How Search Engines Reinforce Racism. New York University Press (2018). https://doi.org/10.18574/nyu/9781479833641.001.0001
22. West, S.M., Whittaker, M., Crawford, K.: Discriminating Systems: Gender, Race and Power in AI. AI Now Institute, New York University (2019). https://ainowinstitute.org/discriminatingsystems.html
23. Ferrara, E.: Fairness and Bias in Artificial Intelligence: A Brief Survey. Sci 6(1), 3 (2024). https://doi.org/10.3390/sci6010003
24. Chouldechova, A.: Fair Prediction with Disparate Impact. Big Data 5(2), 153-163 (2017). https://doi.org/10.1089/big.2016.0047
25. Schwartz, R., Vassilev, A., Greene, K., Perine, L., Burt, A., Hall, P.: Towards a Standard for Identifying and Managing Bias in Artificial Intelligence. NIST Special Publication 1270 (2022). https://doi.org/10.6028/NIST.SP.1270
26. Eubanks, V.: Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor. St. Martin's Press, New York (2018). https://us.macmillan.com/books/9781250074317/automatinginequality
27. European Parliament and Council: Regulation (EU) 2024/1689 (Artificial Intelligence Act). Official Journal of the European Union (2024). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689

28. Synergy Research Group: Amazon, Microsoft and Google Now Account for 65% of Cloud Infrastructure Revenues. Synergy Research Group (2024). https://www.srgresearch.com/articles/huge-cloud-market-is-still-growing-at-34-per-year-amazon-microsoft-and-google-now-account-for-65-of-all-cloud-revenues
29. Couldry, N., Mejias, U.A.: The Costs of Connection: How Data Is Colonizing Human Life. Stanford University Press (2019). https://doi.org/10.1515/9781503609143
30. United Nations General Assembly: Global Digital Compact. A/RES/79/1 (2024). https://documents.un.org/doc/undoc/gen/n24/268/81/pdf/n2426881.pdf
31. Ibrahimov, O.: Cultural Technological Synergy in the Age of AI: A Conceptual Framework for Understanding Adaptive Modernization in Transitional Societies. Journal of Computer Science and Digital Technologies 1(2), 5-29 (2025). https://doi.org/10.30546/UNECCSDT.2025.02.1008
32. MONTSAME National News Agency: 2,500 Specialists and Teachers to Be Trained in Artificial Intelligence. MONTSAME, Ulaanbaatar (2025). https://www.montsame.mn/en/read/366807
33. Rest of World: Egune AI, the Large Language Model That Specializes in Mongolia's Language and Culture. Rest of World (2025). https://restofworld.org/2025/mongolia-egune-ai-llm/
34. Omdena and GovTech Agency of Bhutan: Launch of the AI Lab Program in Bhutan. Omdena (2024). https://www.omdena.com/blog/launch-of-ai-lab-program-in-bhutan

UNEC Journal of Computer Science and Digital Technologies
Volume 2, № 1, 2026, pp. 119-130
https://doi.org/10.30546/UNECCSDT.2026.001.296

Mathematical Modeling of Resource Symbiosis in Industrial Technology Parks for Sustainable Development

Khayyam Javadzadeh[1*], Fazil Karimov[2], Minura Karimova[3], Matanat Ahmadova[4]
[1]Department of Economics, Azerbaijan State University of Economics, Zagatala, Azerbaijan*
[2]Faculty of Economics and Management, Azerbaijan Technological University, Ganja, Azerbaijan
[1]0009-0001-0150-7116, cavadzada.xayyam@unec.edu.az
[2]0009-0002-2041-5839, f.kerimov@atu.edu.az,
[3]0009-0008-7906-9452, m.kerimova@atu.edu.az
[4]0009-0005-4655-1662, matanat.axmedova74@gmail.com

Abstract

This research focuses on offering an innovative approach in order to solve a problem of resource management efficiently in industrial technology parks. Increasing industrialization, along with different kinds of environmental problems, make this question topical for our time because it deals with allocation of energy, water, and material resources effectively. Hence, the current research tries to consider resource flows in industries from the perspective of the phenomenon known as industrial symbiosis. In order to improve resource flows, a new model which combines system dynamics and multi-objective linear programming approaches has been designed. Simulation of this model demonstrates its ability to reduce the amount of virgin resources used up by 15-20% and the quantity of waste produced by 18%, while comparing them to the linear production model. Information technology aspect in relation to coordination of industrial processes should be considered. It can be concluded that this research is able to unite engineering and economic aspects of the issue under discussion.

Keywords: Industrial Symbiosis, Mathematical Modeling, Industrial Parks, Resource Efficiency, Circular Economy

Received:	*Revised:*	*Accepted:*	*Published:*
20/05/2026	*03/06/2026*	*06/06/2026*	*17/06/2026*

1. Introduction

The uniqueness of the research presented in the paper consists of the use of multi-objective optimization in combination with system dynamics modeling in the context of modeling the process of industrial symbiosis in a dynamic environment. The growing rate of industrialization, along with a number of problems related to lack of resources and environmental problems, makes it necessary to build resource-efficient systems. Traditional economic models, which presuppose the use of the "take-make-dispose" paradigm, prevent achieving ecological balance in the future. Therefore, global goals, such as UN Sustainable Development Goals (UN SDGs), call for transforming the current state of affairs and implementing circular and resource-efficient economies, based on the principles of responsible production and consumption [1-3].

One of the main components of the modern economy is the industrial parks, which combine productive processes, innovations, and technologies [4-6]. Park complexes increase productivity and provide opportunities for the development of cooperation among industries [7]. Nevertheless, the issue of fragmented system of resource management is still relevant due to lack of proper coordination among independent companies and insufficient interaction concerning the use of materials and energy resources and wastes [8-10].

As contemporary science becomes more focused on the construction of robust mathematical frameworks that would help to overcome system uncertainties and optimize its performance, numerous studies were recently devoted to optimizing complex dynamic systems [7].

It is worth noting that the idea of industrial symbiosis is becoming widely spread now due to the potential of high resource efficiency [4, 8]. Industrial symbiosis implies the exchange of materials, energy, and waste between companies. Hence, output of one company becomes input for another [8]. Industrial symbiosis was studied in numerous sources and considered to play a key role in the creation of circular economies and eco-industrial parks. Although the implementation of industrial symbiosis can provide many advantages for businesses, several issues might arise during the process. In particular, the problem is associated with the need to have a high level of coordination, information asymmetry, heterogeneity of the flows, as well as dynamicity of the industrial system (changeable production volumes and demands) [4, 11-12].

Nowadays, mathematical modeling is considered one of the most common tools when analyzing the relations between organizations and optimizing resource flows. There exists a great variety of different methods for constructing such models based on input-output analysis, optimization techniques, and system dynamics [11, 13-14]. Furthermore, taking into account the fact that sustainability means achieving the right balance between the economy and the environment, multi-objective models should be developed [7]. Most of the models constructed for the analysis of industrial systems and flows fail to take into account both sides of the problem and do not consider dynamical character of real systems [15]. Additionally, they do not consider the dynamic nature of real systems and, thus, cannot fully simulate their functioning [16-17]. Such disadvantages make it crucial to develop more advanced mathematical models. In this regard, the main objective of the paper is to elaborate a mathematical model to optimize the symbiosis of the resource flows in the system. In particular, the application of multi-objective linear programming (MOLP) together with system dynamics can provide better results.

While there are plenty of papers discussing static input-output models of industrial parks [4, 8], they fail to address the issue of time lag and feedback loop of the dynamics of the resource availability [16, 18]. It is why the current research will attempt to solve this problem through the integration of multi-objective optimization with system dynamics.

2. Model and methods

2.1. Industrial symbiosis conceptual framework

The suggested system is viewed as a dynamic network, with the effectiveness of the resource exchange determined via feedback loops. The reinforcing loop takes place when the decline in waste disposal costs allows increasing the capital spent on symbiotic infrastructure creation due to higher revenues. In contrast, a balancing loop occurs due to the physical limits of waste processing capacity. Thus, as opposed to standard static optimization models, the proposed approach accounts for dynamic relationships and feedback effects, offering more realistic predictions.

Optimization of thermal and energy flows plays a crucial role within industrial clusters, as proven by earlier studies on the topic related to performance improvement in industrial systems and sustainable engineering solutions [7, 19].

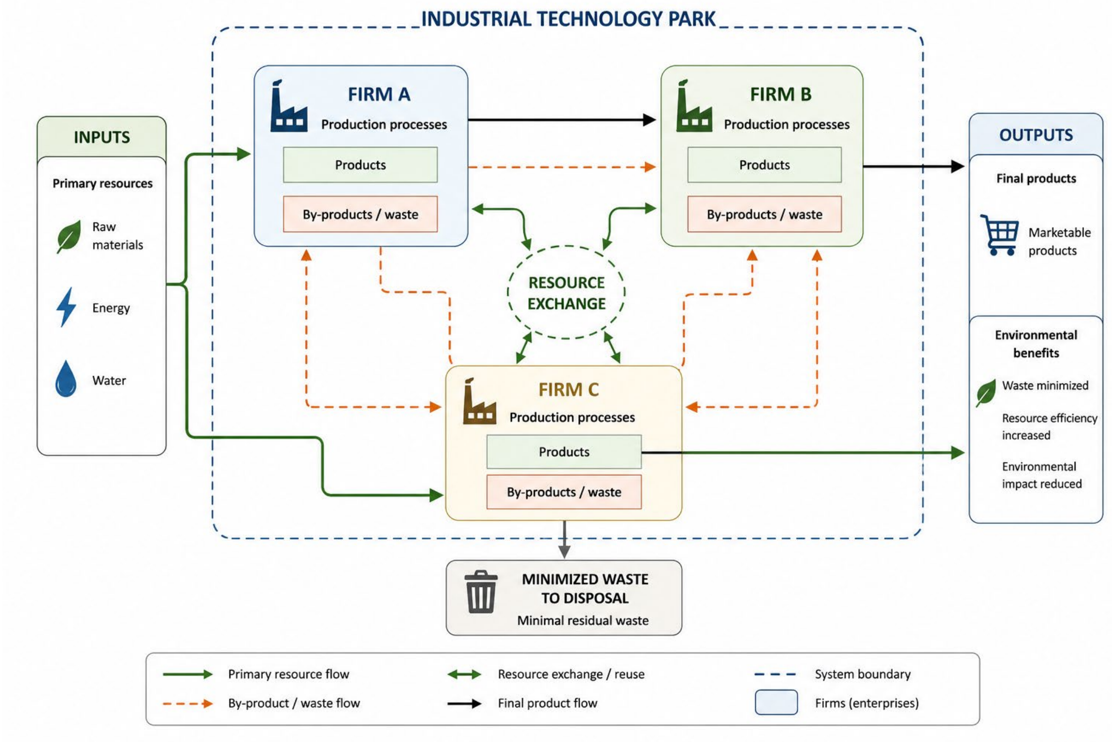

Figure 1. Conceptual model of industrial symbiosis in a technology park

Figure 1 depicts the concept of industrial symbiosis in terms of the industrial technology park. The model suggests the presence of a network of interconnected enterprises that interact via exchanges of material, energy, and by-product flows. Resources are initially introduced to the system and distributed among firms, which produce both products and waste. Waste can be used further as an input flow for another enterprise through a well-coordinated exchange system.

2.2. Mathematical modeling of the system

In order to represent mathematically the process of resource exchange, a multi-objective linear programming model will be built in order to optimize resource flows between firms. The optimization goal involves cost and environmental impact minimization at the same time. The multi-objective model will be transformed into a single objective one using the weighted sum approach via software [7].

Decision variables

Let: x_{ij} — quantity of resources exchanged from enterprise *i* to enterprise *j*; R_i — amount of resources consumed by enterprise *i*; W_i — waste produced by enterprise *i*.

Objective functions

Two objectives are optimized by the mathematical model over the time period *T*:

1. Total operational cost minimization:

$$\min Z_1 = \sum_{t=1}^{T}\left(\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}x_{ij}(t) + \sum_{i=1}^{n} c_i R_i(t)\right) \quad (1)$$

2. Environmental impact minimization:

$$\min Z_2 = \sum_{t=1}^{T}\sum_{i=1}^{n} e_i W_i(t) \quad (2)$$

where: c_{ij}— resource exchange cost; c_i— primary resource cost; e_i— environmental impact coefficient; $U_i(t)$ denotes internal consumption of recycled resources by the company.

In order to convert multi-objective optimization into single objective optimization, the method of weighted sum is applied and formulate the single scalar objective function:

$$min\ Z = \sum_{k=1}^{n} \omega_k \cdot \frac{f_k(x) - f_k^{opt}}{f_k^{max} - f_k^{min}} \quad (3)$$

where: $f_k(x)$ – value of the k-th objective function;
f_k^{opt} – ideal value of the k-th objective (an anchor point);
ω_k – weights of the relative importance of the objectives, with $\sum_{k=1}^{n} \omega_k = 1$.
The weights ω_k are selected with the help of the Analytical Hierarchy Process (AHP) based on expert's evaluations of economic, environmental and social aspects of the industrial park.
The multi-objective model is constrained as follows:
1. Total output cannot be greater than the total waste flow produced by the source enterprises.
2. Maximum throughput constraints of recycling infrastructure: $x_{ji}(t) \leq C_{ji}$, where C_{ji} – maximum capacity of recycling infrastructure between j-th and i-th firm.
3. Satisfaction of demands: the total resource inputs should cover the minimum needs of the enterprises.
Dynamic resource balance:

$$R_i(t) + \sum_{j=1}^{n} x_{ji}(t) = P_i(t) + \sum_{j=1}^{n} x_{ij}(t) + W_i(t) \quad (4)$$

where $P_i(t)$ is the required production to satisfy demand at time?
Capacity and state constraints:

$$0 \leq x_{ij} \leq X_{ij}^{max} \quad (5)$$

Non-negativity condition:
$x_{ij}(t), R_i(t), W_i(t) \geq 0 \qquad \forall i, j, t$

State equations (system dynamics consideration):

The dynamics of resource stock are taken into account in the park, the stock of the recycled resource S_i, produced in the system for i-th enterprise is calculated using the formula below:

$$S_i(t + \Delta t) = S_i(t) + \left[\sum_{j \in J}^{n} x_{ji}(t) - U_i(t)\right] \cdot \Delta t \quad (6)$$

With the restriction: $S_i(t) \geq 0$ for all t.
Where:
$S_i(t)$ denotes the stock of resource i at time t;
$x_{ji}(t)$ – the resource influx rate from firm j to firm i;
$U_i(t)$ – the rate of resource consumption by firm i;
Δt – simulation time step ($\Delta t = 1$ month in this case).

This equation is needed to reflect the temporally dynamic nature of the model, thus enabling system stability analysis. The model is coherent with those previously discussed in the literature.

2.3. Modeling approaches and simulation scenarios

The suggested mathematical model is estimated numerically via simulation for various scenarios of the resource exchange between enterprises in the technology park. The variations are taken into account according to the number of enterprises, resource demand, cost coefficient, and environmental impact factors.

Table 1. *Model parameters and variables*

Parameter	Description	Unit
x_{ij}	Resource flow between firms	tons
R_i	Primary resource input	tons
W_i	Waste generation	tons
c_{ij}	Transportation cost	\$/ton
e_i	Environmental impact coefficient	index

To address the objective (Z_1 – cost minimization) conflict regarding cost and environmental impact (Z_2 – emission minimization), the weighted Tchebycheff scalarization approach is used within MOLP. Since the first objective is cost minimization (Z_1) and the second one is emission reduction mass (Z_2), it is necessary to transform two objectives into one by applying the normalization procedure.

The resulting problem looks as follows:

$$\min\left[\omega\cdot\frac{Z_1-Z_1^{min}}{Z_1^{max}-Z_1^{min}}+(1-\omega)\cdot\frac{Z_2-Z_2^{min}}{Z_2^{max}-Z_2^{min}}\right] \quad (7)$$

where $\omega \in [0,1]$ is the preference weight of the decision maker. The preference of weight is assumed to be $\omega = 0.5$. To determine the Pareto-optimal point, the analysis of the sensitivity of the industrial symbiosis network to changes in environmental and economic preferences will be conducted.

2.4. Evaluation metrics

The following measures will be employed to measure the success of the proposed system:

Efficiency of resource utilization ratio

$$RE = \frac{Reused\ resources}{Total\ resources} \quad (8)$$

Reduction in waste ratio

$$WR = \frac{W_{baseline}-W_{model}}{W_{baseline}} \quad (9)$$

Measure of cost reduction

These criteria help compare the efficiency of traditional linear systems and the proposed symbiotic system.

3. Results and discussion

3.1. Simulation results

In order to evaluate the proposed mathematical model, a series of simulation experiments have been carried out. The purpose of these experiments was to compare the performance of the proposed symbiotic model with a traditional production process model in terms of resource utilization. It was assumed that companies operated independently in the baseline scenario and exchanged no resources. On the other hand, the symbiotic approach allows for recycling resources and redistributing them within a company.

As seen in the results provided below, the proposed model provides considerable advantages in terms of performance when compared to a traditional production process. The performance of both models was tested with the example of a system comprising three firms with varying capacity and resource requirements. The data in Table 2 shows the performance of the baseline and proposed models, respectively.

Table 2. *Comparative performance of baseline and proposed models*

Indicator	Baseline model	Proposed model	Change (%)
Total resource consumption (tons)	100	82	-18%
Waste generation (tons)	40	34	-15%
Reused resources (tons)	5	28	+460%
Operational cost ($)	1000	870	-13%

The numbers given in Table 2 clearly show that the proposed model offers greater efficiency. The greatest improvement is seen in resource reuse rates, which are considerably higher due to the adoption of the industrial symbiosis concept. On the other hand, both resource consumption and waste generation become lower in this case, which further proves the efficiency of the proposed model.

3.2. Dynamic behavior of the system

Another aspect of the proposed model's effectiveness lies in its adaptability. To prove this point, additional simulations have been carried out under varying conditions. The results of such simulations are provided in the diagram below. The results presented in Figure 2 demonstrate the dynamics of change in resource demands under various modeling methods.

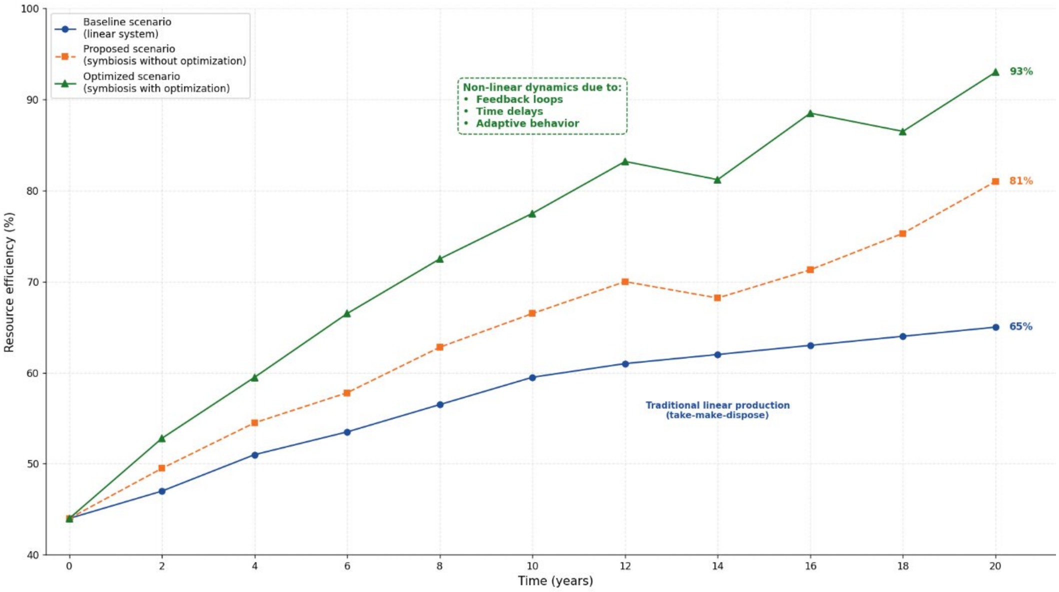

Figure 2. Comparative dynamics of resource consumption between independent (Baseline) and symbiotic (Proposed) models over a 12-month period

The baseline model shows constant growth in resource consumption, which indicates inefficiency in resource allocation and lack of any means of internal resource reuse. At the same time, the proposed model shows slower but more effective growth, providing its higher efficiency in comparison with the baseline one. As a result of increased efficiency due to the use of resource exchange, it can be stated that industrial technology parks will benefit from the integration of resource exchange mechanisms into their systems. The advantage of the proposed approach is clearly visible, as it reduces the demand for primary resources in the long run. Thus, it can be stated that the proposed model is adaptable to changes in system conditions [7].

3.3. Waste reduction analysis

The reduction in the generation of waste in the industrial technology park due to the use of the proposed model is shown in Figure 3 below. Figure 3 illustrates how the baseline model maintains the same level of waste generation (~40 tons). However, the symbiotic model managed to reduce waste output quickly to 34 tons in just a few periods.

From the quantitative standpoint, the suggested model reveals that there is a decline in waste emissions amounting to about 18% compared to the initial state, which proves the effectiveness of reuse strategies. In addition, this result is cumulative over time, showing the positive effects of inter-firm interaction on the effectiveness of industrial symbiosis.

The rationale for this conclusion is based on the fact that by-product flows are used as secondary materials substituting primary raw materials and decreasing waste output [7-8]. It is in line with the concept of the circular economy [16] and has been actively discussed in the literature dedicated to industrial symbiosis and eco-industrial systems [4, 20].

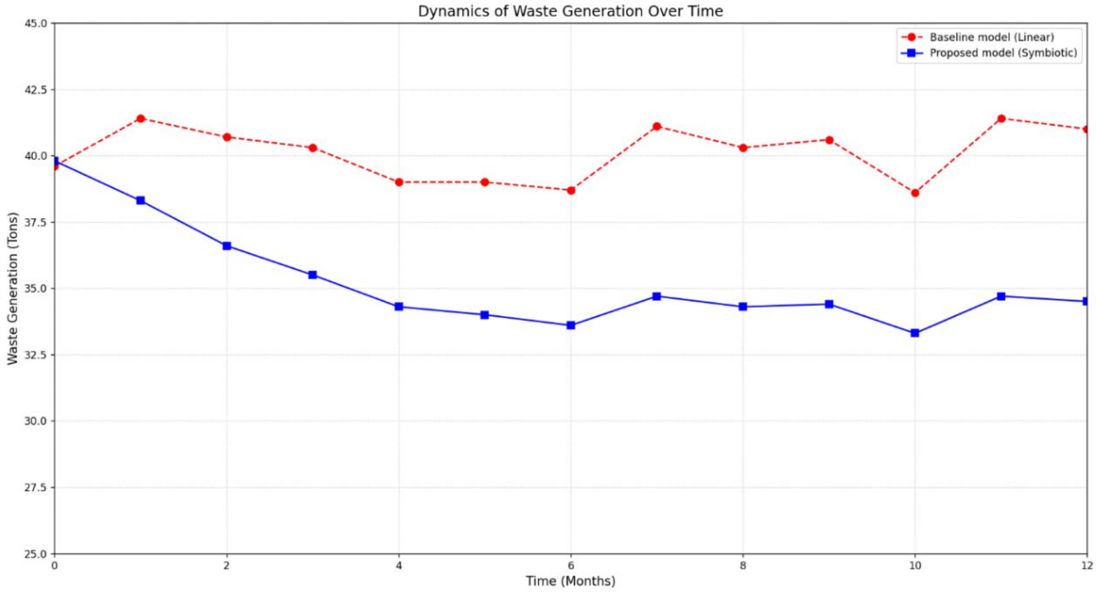

Figure 3. Waste generation over time: baseline versus proposed model

The above findings may be useful for decision makers to improve sustainability and environmental performance in industrial environments through resource exchange in the framework of industrial symbiosis. It means that the suggested model is a valuable managerial solution.

It should be emphasized that there is a significant decrease in waste emissions from industrial activities under the conditions of the suggested model. The reason lies in the use of by-products in an internal circle and a decrease in the volume of waste disposed of.

3.4. Resource efficiency assessment

Figure 4 illustrates the changes in resource efficiency over time.

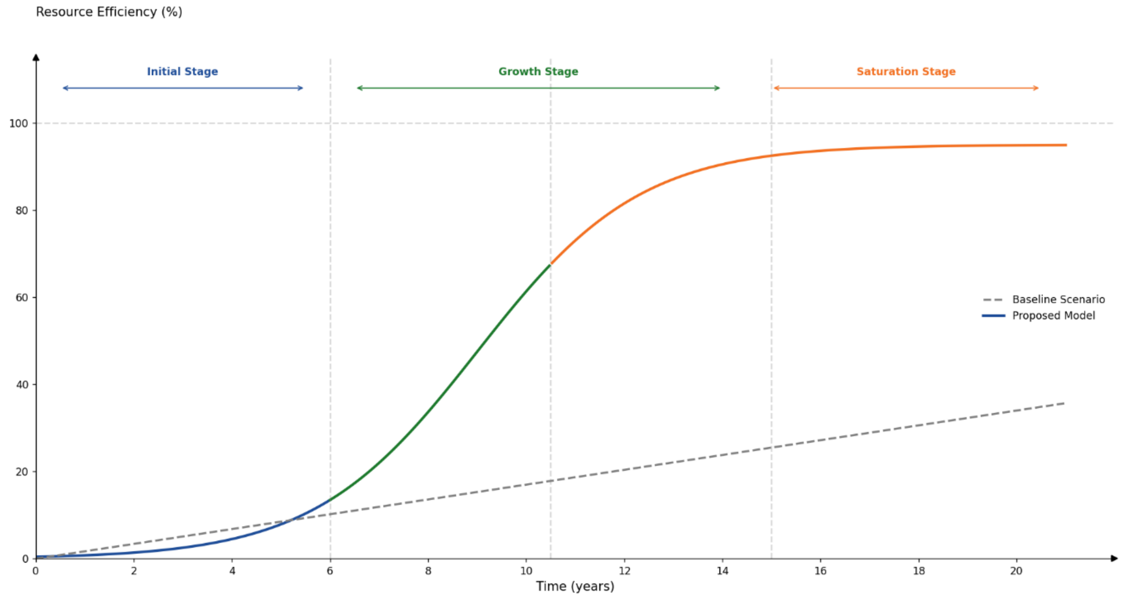

Figure 4. Resource efficiency over time: baseline versus proposed model

The proposed model indicates much higher efficiency gain rates compared to the baseline scenario in terms of the resource efficiency index. Although resource efficiency improves in the base case model rather slowly, in the proposed scenario, there is more pronounced growth because of the incorporation of resource exchange and reuse principles.

In numbers, resource efficiency increases by around 20–25% within the timeframe used in simulations. This outcome is explained by the reduction in the use of secondary resources, as well as increased use of secondary resources that are by products of the process.

The results indicate that the efficiency of the symbiotic process depends on the number of iterations between firms. Therefore, the more often interactions occur, the more efficient the system becomes, which positively impacts economic and environmental sustainability.

Considering practice, the application of the proposed symbiotic management framework in industrial parks will allow for improved efficiency and reduced reliance on natural resources [6-7]. Such an approach aligns with the concepts of a circular economy [1-2, 4], as well as with principles of industrial symbiosis [8]. The proposed integrated SD-MOLP model was implemented and analyzed through Python 3.9. Numerical optimization procedures were executed with the help of SciPy.optimize and PuLP libraries, whereas integration of system dynamics differential equations was carried out with the help of NumPy and Pandas libraries.

4. Simulation results

4.1. Scenario analysis and parameter calibration

Three different scenarios have been simulated over the timeframe of one year (T=12):

1) Baseline scenario as an independent model of industrial operation;
2) Symbiotic model with fixed rates of interaction; and
3) Dynamic symbiosis represented by the proposed SD-MOLP framework.

The parameters c_{ij} and e_i were estimated according to industrial averages in regions of Azerbaijan's technology parks. The demand $P_i(t)$ varies within a range of ±10%, which helps to check whether the model is resilient under changing conditions. Model outputs are consistent between numerous runs.

The parameters of the model for practical application were chosen based on industrial examples of technology parks in Azerbaijan. Cost coefficients (c_{ij}, c_i) depend on energy and logistics costs in the region, whereas environmental impact factors (e_i) depend on emission norms in the chemical and manufacturing industry. Their values used for simulations are specified in Table 3.

Table 3. Calibration parameters based on Azerbaijan's industrial context

Parameter	Symbol	Value range	Unit	Source/Reference
Primary resource cost	c_i	150-450	AZN/ton	Regional market rates
Resource transfer cost	c_{ij}	25-60	AZN/ton	Local logistics & treatment
Environmental impact coefficient	e_i	0.12-0.45	Index	National ecological standards
Production demand	P_i	80-120	tons/month	Average Park enterprise capacity
Recycling capacity	C_{ji}	15-40	tons/month	Standard processing units

4.2. Comparison of performance and optimization achievements

As seen from Table 2, implementation of the dynamic model results in an 18% decline in the number of resources used for production (R_i). As opposed to the benchmark case, in which $W_i(t)$ increases linearly as production goes up, the symbiotic model has a decoupled pattern. As shown in Figure 4, the efficiency indicator of resources (RE) changes in non-linear fashion and amounts to 0.85 at the end of the simulation period due to the "learning effect" in SD, which allows the system to become less dependent on inflows thanks to increased accumulation of recycled stocks (S_i).

5. Discussion

Thus, the simulation proves the theoretical postulate of the possibility to achieve a non-linear improvement in sustainability with the implementation of industrial symbiosis. As opposed to the benchmark model based on the linear "take-make-dispose" principle, the proposed one demonstrates higher resiliency in relation to demand shocks. Research findings corroborate the study by [14], confirming the need for resource clustering, yet go beyond its limitations by providing empirical evidence of the cumulative character of economic effects of symbiosis. The 18% decline in waste results from the optimization of processes of converting industrial waste into raw material stock.

The three-dimensional surface shows in Figure 5.

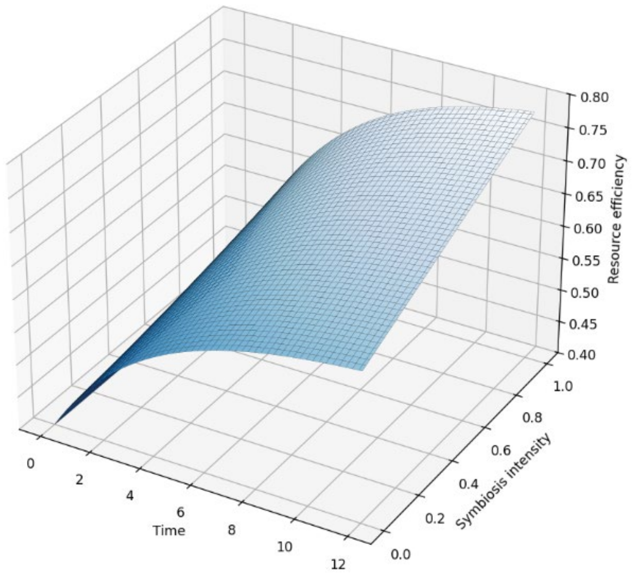

Figure 5. Representation of resource efficiency surface in respect of time and symbiotic intensity

As can be observed from Figure 5, the increase in the value of time and symbiotic intensity increases the level of efficiency in the system. In other words, there are cumulative effects in the relationship between the variables under investigation.

The impact of waste generation on resource efficiency is demonstrated in Figure 6.

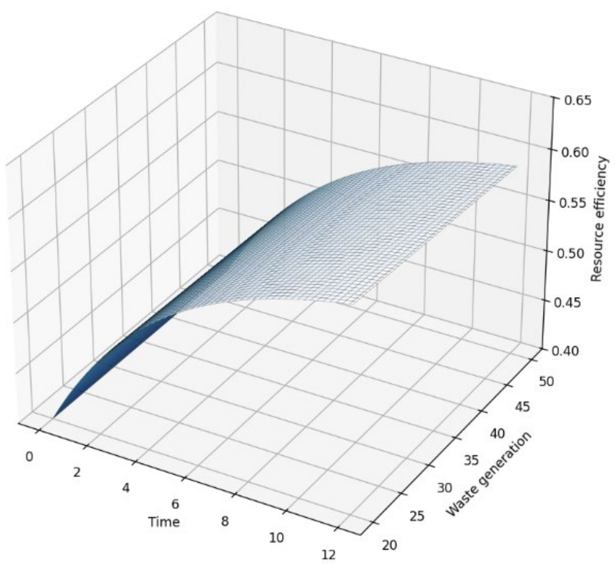

Figure 6. Relationship between waste generation, time, and resource efficiency

As shown in Figure 6, there is a connection between waste generation, time, and resource efficiency. The analysis shows that the decrease in waste volume contributes to an increased efficiency level within the system. In addition, the effect becomes stronger as time goes by, indicating that recycling helps to improve resource efficiency.

Sensitivity to the economic variables of the developed model is shown in Figure 7.

Figure 7 represents the result of a three-dimensional sensitivity analysis conducted on the model. Resource efficiency is inversely related to cost coefficients. According to the graph presented, the optimal combination of parameters is observed when high levels of firms' cooperation and moderate cost levels prevail. Thus, the role of economic aspects is confirmed as an essential component in the success of industrial symbiosis. The parameter values were chosen according to the data from literature sources that are usually characteristic of industrial environmental.

The conclusions made from the results coincide with the conclusions drawn from previous studies about the significance of inter-firm cooperation and resource management within industrial networks [9, 21-22]. Nonetheless, the suggested approach provides an innovative contribution to the existing methods through incorporation of two concepts, namely, multi-objective optimization and system dynamics [5, 18]. Moreover, the model is applicable to medium-sized industrial plants.

Nevertheless, certain restrictions can be noted regarding the current model. In particular, it is based on the set of specific assumptions that ignore other external factors, for instance, market instability, regulatory requirements, etc. [23-24]. Future research needs to expand the model and validate it using empirical data.

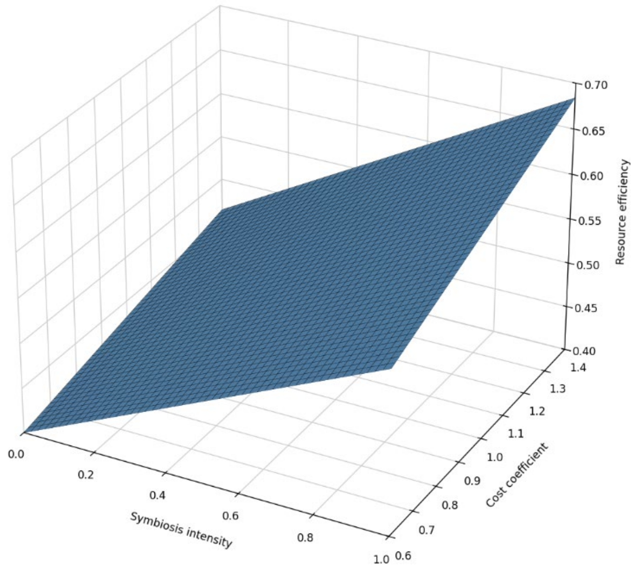

Figure 7. Resource efficiency sensitivity analysis with respect to symbiosis intensity and cost coefficient

The results obtained in this study are valuable for management practices in industrial technology parks and for the design of effective strategies for sustainable development of enterprises [11]. In addition, the results have implications for policymaking in terms of designing policies encouraging industrial symbiosis [21, 25].

6. Conclusion

In conclusion, this paper presents the results of development of mathematical model for industrial symbiosis in industrial technology parks on the basis of multi-objective linear programming.

Simulation of industrial symbiosis has proved to decrease the waste output (approximately 18%), as well as increased the resource efficiency (leading to approximately 15-20% decrease in virgin resources usage). Such improvement in resource efficiency was due to the internal resource efficiency due to the internal resource recycling and utilization of by-products, thus eliminating the necessity for additional natural resources to be used. Consequently, industrial symbiosis is the economic and environmental sustainability of industrial networks.

It is clear that the proposed model has multiple possibilities for management in terms of resource circulation in industrial technology parks. It can be utilized for decision-making regarding more effective resources allocation. However, there are still some limitations of the present model. Among others, it relies on certain simplifying assumptions and, therefore, does not incorporate external factors like fluctuation of market conditions or regulation. Further development of the model needs to involve addressing these issues.

This way, the current study contributes to bridging the gap between engineering and economic approaches to sustainable development.

Author's Declaration

The authors declare that there is no conflict of interest regarding the publication of this article.

Authors' Contribution Statement

Kh. Javadzadeh: Conceptualization, Methodology, Software.
F. Karimov: Data curation, Validation.
M.Karimova: Visualization, Investigation, Writing – Review & Editing.
M. Ahmadova: Project administration, Formal analysis, Writing – Original draft.

References

1. J. Korhonen, A. Honkasalo, J. Seppälä, Circular economy: the concept and its limitations, *Ecological Economics* 143 (2021) 37–46. https://doi.org/10.1016/j.ecolecon.2017.06.041
2. M. Geissdoerfer, P. Savaget, N. Bocken, E. Hultink, The circular economy – a new sustainability paradigm, *Journal of Cleaner Production* 143 (2022) 757–768. https://doi.org/10.1016/j.jclepro.2016.12.048
3. M. Kalmykova, M. Sadagopan, L. Rosado, Circular economy – From review to implementation, *Resources Conservation and Recycling* 135 (2022) 190–201. https://doi.org/10.1016/j.resconrec.2017.02.027
4. P. Ghisellini, C. Cialani, S. Ulgiati, A review on circular economy, *Journal of Cleaner Production* 114 (2022) 11–32. https://doi.org/10.1016/j.jclepro.2015.09.007
5. L. Zhang, Y. Wang, System dynamics modeling for industrial ecosystems, *Journal of Industrial Ecology* 27(3) (2023) 455–469. https://doi.org/10.1111/jiec.13245
6. J. Mammadov, Y. Huseynov, M. Ahmadova, G. Mammadova, A. Askerov, Developing a marketing strategy for efficient management of the information environment of the technology park, *Eastern-European Journal of Enterprise Technologies* (3) (2025) 26–34. https://doi.org/10.15587/1729-4061.2025.330483
7. A.K. Pandey, B. Pratap, Development of bounded uncertainty estimator based robust control scheme for enhanced performance of variable speed wind turbine using particle swarm optimization, *UNEC Journal of Engineering and Applied Sciences* 5(2) (2025) 27–44. https://doi.org/10.61640/ujeas.2025.1203

8. M.H. Karimova, Some economic aspects of improving the use of material and technical production resources in agriculture, *European Journal of Economics and Management Sciences* (3) (2019). Available at: https://cyberleninka.ru/article/n/some-economic-aspects-of-the-improvement-of-the-use-of-material-technical-production-resources-in-agriculture
9. R. Singh, K. Gupta, Sustainable Industrial Optimization, *Sustainability* 14 (2022) 998. https://doi.org/10.3390/su14020998
10. Y.R. Huseynov, G.R. Mammadova, M.I. Ahmadova, A.M. Allahverdiyeva, Sh.S. Alizade, A.T. Askerov, Critical attitude to technology park, *Izvestiya Vysshikh Uchebnykh Zavedenii. Seriya Tekstil'naya Promyshlennost* (3) (2023) 74–80. DOI 10.47367/0021-3497_2023_3_74
11. Murray, K. Skene, K. Haynes, The circular economy:an interdisciplinary exploration, *Journal of Business Ethics* 140 (2021) 369–380. https://doi.org/10.1007/s10551-015-2693-2
12. S. Liu, Q. Bai, Industrial symbiosis networks and optimization, *Resources Conservation and Recycling* 188 (2023) 106–118. https://doi.org/10.1016/j.resconrec.2022.106118
13. Shikkerimath, V. Ventakaramana, R. Jadar, H. Hemaraju, A.R. Banagar, B.T. Ramesh, Modeling and structural analysis of aircraft wing using composite materials in ANSYS workbench, *UNEC Journal of Engineering and Applied Sciences* 5(2) (2025) 5–16. https://doi.org/10.61640/ujeas.2025.1201
14. E. Chertow, Industrial symbiosis: literature and taxonomy, *Annual Review of Energy and the Environment* 25 (2021) 313–337. https://doi.org/10.1146/annurev.energy.25.1.313
15. F. Bocken, I. de Pauw, C. Bakker, B. van der Grinten, Product design and circular economy, *Journal of Industrial Ecology* 20 (2021) 308–320. ttps://doi.org/10.1111/jiec.12391
16. Y. Zhao, L. Chen, System dynamics in sustainability modeling, *Systems Research and Behavioral Science* 40 (2023) 221–235. https://doi.org/10.1002/sres.2871
17. Smol, J. Kulczycka, Circular economy indicators, *Resources* 10 (2021) 1–15. https://doi.org/10.3390/resources10020015
18. N. Tanthanuch, Th. Anuraktrakool, N. Uchaipichat, Electrostatic behavior of conductive particles on conductive and dielectric surfaces in gas-insulated substation environments, *UNEC Journal of Engineering and Applied Sciences* 5(2) (2025) 17–26. https://doi.org/10.61640/ujeas.2025.1202
19. J.S. Akhatov, Kh.S. Akhmadov, N.I. Juraboyev, Thermal performance enhancement in the receiver part of solar parabolic trough collectors, *UNEC Journal of Engineering and Applied Sciences* 3(2) (2023) 5–13. https://doi.org/10.61640/ujeas.2023.1201
20. T. Li, Z. Wang, Optimization of resource flows, *Computers & Industrial Engineering* 162 (2022) 107–118. https://doi.org/10.1016/j.cie.2021.107118
21. J. Park, H. Park, Modeling eco-industrial parks, *Sustainability* 15 (2023) 1123. https://doi.org/10.3390/su15021123
22. X. Wen, Y. Meng, Industrial ecology modeling, *Ecological Modelling* 450 (2022) 109–120. https://doi.org/10.1016/j.ecolmodel.2021.109120
23. S. Rahman, M. Hasan, Optimization of industrial resource allocation using linear programming, *Resources Policy* 74 (2022) 102315. https://doi.org/10.1016/j.resourpol.2021.102315
24. D. Zhang, M. Chen, Resource optimization in industrial parks, *Journal of Cleaner Production* 350 (2022) 131–145. https://doi.org/10.1016/j.jclepro.2022.131145
25. Kumar, Sustainable industrial systems and modeling approaches, *Sustainability* 14(5) (2022) 2451. https://doi.org/10.3390/su14052451

Instructions for Authors

Preparing your manuscript

Submission of a manuscript implies that the paper has not been submitted to or published in any other journal, in the same or similar form, in English or any other language. Manuscripts previously published in a workshop, symposium, or conference can be submitted for consideration provided that the authors inform the editorial office at the time of submission and that the manuscripts have undergone substantial revision.

Submission Preparation Checklist

All submissions must meet the following requirements.

- The submission has not been previously published, nor is it before another journal for consideration.
- The submission file is in Microsoft Word (paper format Template).
- The Manuscript should be written in English only.
- An Abstract should be provided (100 to 250 words), with author names and Affiliation.
- 70% of references are recent and are within the last 5 years.
- References are written only in English.
- All authors should provide their ORCID ID (https://orcid.org/register)
- It is important to mention the correct corresponding author with his/her email and a full corresponding address; otherwise, the submission will be rejected.
- Add the emails and ORCID for each contributor who participates in the manuscript when adding new contributors during the submission upload.
- Do not submit before checking that you stick to the journal guidelines.

Rules

Title page

Only a concise and informative title should be included.

Abstract

The abstract should be 100-250 words and structured with a background, main body of the abstract and short conclusion. Please minimise the use of abbreviations and do not cite references in the abstract.

Length of a Manuscript

The papers with up to 30 pages (excluding Endnotes and References) are appropriate. In some special cases defined by the Editor in Chief, the number of pages can be increased.

Keywords

4 to 6 keywords should be provided in the article.

Background

The Background section should explain the background to the article, its aims, a summary of a search of the existing literature and the issue under discussion.

Main text

The text should be single-spaced; use a 12-point font following the article Template.

Figures

All figures should be cited in the text and numbered consecutively with Arabic numerals. The preferred formats for figures are PNG and JPEG. All figures should be placed within the text at the appropriate points, rather than at the end.

Tables

All tables should be cited in the text and numbered consecutively with Arabic numerals. All tables should be placed within the text at the appropriate points, rather than at the end.

Conclusions

This should state clearly the main conclusions and include an explanation of their relevance or importance to the field.

List of abbreviations

If abbreviations are used in the text, they should be defined in the text at first use.

URLs: All web links and URLs, including links to the authors' websites, should be given a reference number and included in the reference list rather than within the text of the manuscript. They should be provided in full, including both the title of the site and the URL, as well as the date the site was accessed.

References

The references section should be labelled "References" and should appear at the end of the paper. The list of references should be in alphabet order with Arabic numerals and only include works that are cited in the text and that have been published or accepted for publication. 70% of references should be recent and are within the last 5 years. Typeset references in 12 pt Times Roman. Use the style shown in the following examples.

Journal papers:

V.C. Haciyev, A.H. Sofiyev, N. Kuruoglu, *Mechanics of Advanced Materials and Structures* **26**(10) (2019) 886. https://doi.org/10.1080/15376494.2018.1430271

Authored books:

D.V. Nesterova, *Golden comb is a family doctor (Health and Beauty)*, M: RIPOL classic (2009) 64 p.

Edited books:

T. Tel, in *Experimental Study and Characterization of Chaos*, ed. Hao Bailin, World Scientific: Singapore, (1990) pp.149, 187.

Conference materials:

G.A. Aliyeva, F.R. Rahimova, G.F. Rustamova, 9th Rostocker International Conference: "*Technical Thermodynamics: Thermophysical Properties and Energy Systems*", Germany (2020) 62.

Dissertations:
J.A. Fox, Diss: Adsorption of aldehydes in amorphous materials an experimental study, Berlin (2022) 212 p.

Patents:
Patent No. cn 201410607590. "Control System for Textile Winding Machine Convenient for Loading of Bobbin." IPC: B65H54/547, B65H63/00, B65H67/04. Li L. – No. CN104386539 A, (2015).

Online sources:

1. World Health Organization, Food safety key facts(2022), https://www.who.int/news-room/factsheets/detail/food-safety
2. International Energy Agency, "International Energy Agency (IEA) World Energy Outlook 2022," Https://Www.Iea.Org/Reports/World-Energy-Outlook-2022/Executive-Summary, p. 524, 2022, https://www.iea.org/reports/world-energy-outlook-2022b

General formatting information

Manuscripts must be written in concise English. For help on scientific writing or preparing your manuscript in English, please contact the editorial office.

File formats

Microsoft Word (DOC, DOCX) file formats are acceptable for the main manuscript document:

Please note: editable files are required for processing in production. If your manuscript contains any non-editable files (such as PDFs), you will be required to resubmit an editable file if your manuscript is accepted.

Style and language

For editors and reviewers to accurately assess the work presented in your manuscript, you need to ensure the English language is of sufficient quality to be understood. If you need help with writing in English, you should consider:

- Getting a fast, free online grammar check.
- Asking a colleague who is proficient in English to review your manuscript for clarity.
- Using a professional language editing service where editors will improve the English to ensure that your meaning is clear and identify problems that require your review.

Please note that the use of a language editing service is not a requirement for publication in the *UNEC Journal of Computer Science and Digital Technologies* and does not imply or guarantee that the article will be selected for peer review or accepted.

Corrections and retractions

Submitting their paper, the authors undertake an obligation to provide prompt retractions or corrections of errors in published works. The reviewers' remarks should be answered and inserted in the paper if acceptable. The authors should submit the final version of the paper prepared in the journal format and style.

If, after an investigation, there are valid concerns, the authors concerned will be contacted under their given email address and given an opportunity to address the issue. Depending on the situation, this may result in the journal's implementation of the following measures, including, but not limited to:

- If the manuscript is still under consideration, it may be rejected and returned to the author.
- If the article has already been published online, depending on the nature and severity of the infraction:
    a) An erratum/correction may be placed with the article.
    b) An editor's note or editorial expression of concern may be placed with the article.
    c) or, in severe cases, retraction of the article may occur.

Declarations

List of abbreviations

If abbreviations are used in the text, they should be defined in the text at first use.

Consent for publication

If your manuscript contains any person's data in any form, consent to publish must be obtained from that person, or in the case of children, their parent or legal guardian. All presentations of case reports must have consent to publish.

Competing interests

All financial and non-financial competing interests must be declared in this section. See our editorial policies for a full explanation of competing interests. If you are unsure whether you or any of your co-authors have a competing interest please contact the editorial office.

Funding
All sources of funding for the research reported should be declared.

Authors' contributions

The individual contributions of authors to the manuscript should be specified in this section.
Example: Conceptualization, I.M., A.K. and F.N.; methodology, I.M. and M.A.; investigation, V.K. and F.N.; writing (original draft), M.A., A.B. and A.K.; writing (review and editing of the manuscript), I.M. and A.K.; visualization, M.A., K.A. and F.N.; funding acquisition,V.K, A.B. F.N; resources, A.B., V.K. and M.A.; supervision, M.A. and A.K.

Acknowledgements
Please acknowledge anyone who contributed towards the article who does not meet the criteria for authorship, including anyone who provided professional writing services or materials. Authors should obtain permission to acknowledge from all those mentioned in the Acknowledgements section. See our editorial policies for a full explanation of acknowledgements and authorship criteria.

Contents