



JOURNAL OF SOFTWARE ENGINEERING AND APPLIED SCIENCES

Volume 1, Issue 1; ISSN: 0000-0000



EDITORIAL BOARD OF THE JOURNAL:

A.V. Umarov (chairman), E.B. Saitov (deputy chairman), T. A. Khujakulov

Editor-in-Chief

T. A. Khujakulov

Deputy Editor-in-Chief

E.B. Saitov

Secretary-in-Chief

A. Sadenov

Technical Editors:

N.N. Azimov

Computer Science and Software Engineering:

Scholars of the Republic

S.Babayev, Sh Turaev, A.Narzullayev,

Foreign Scholars

Andrew Ng, Fei-Fei Li, Justin Yifu Lin, Michael

I. Jordan, Yoshua Bengio

Multimedia Technologies, Pedagogy and Psychology:

Scholars of the Republic

Dilnoza Saidova, Gulnora Ziyayeva, Shavkat Juraev.

Foreign Scholars

Richard E. Mayer, John Sweller, Clark Quinn, Daniel Suthers.

Mathematical Modeling and Computational Methods:

Scholars of the Republic

Bakhodir Sadykov, Anvarbek Abdullayev.

When reprinting materials, it is necessary to indicate that they were taken from "Social Sciences: Theory and Practice" (<https://science-utas.org/index.php/ppef/en>)

Foreign Scholars

Gilbert Strang, Lloyd N. Trefethen, Alfio Quarteroni, J. Tinsley Oden

Applied Mathematics and Informatics:

Scholars of the Republic

Ingrid Daubechies, Tursunboy Rakhimov, Dilshod Jurayev

Foreign Scholars

Terence Tao, Ingrid Daubechies, Peter Deuflhard, Nick Higham.

Information Technologies in Education:

Scholars of the Republic

Sherzod Turaev, Akmal Saidov, Dilnoza Saidova.

Foreign Scholars

Curt Bonk, Diana Laurillard, Tony Bates, Neil Selwyn

The authors are responsible for the accuracy of the materials and evidence in advertisements published in the journal. The editorial opinion may not always coincide with the opinion of the authors. Submitted manuscripts will not be returned.

Our address: 100149, Tashkent – 95, Gavkhar Street, №1A. Tel.: 71200 05 40, e-mail: info@utas.uz

CONTENT

1. Rasa Theory and Digital News: Mapping Bharata’s Natyashastra as Engagement Metric Onto Click-Through Rates In Digital News Reels	4
<i>Tasha Singh Parihar, Pramod Kumar Pandey</i>	
2. Hydrogeological and Engineering-Geological Characteristics of the Handise Field ..	18
<i>Khujakulov Toshtemir, Majidova Yulduz, Choriye Anvar</i>	
3. Governance-as-Code for Deterministic, Reliable and Reproducible AIoT Systems in Smart Buildings	22
<i>Young Im Cho, Saydirasulov Norkobil, Davronbekov Dilmurod, Dusmurod Kilichev</i>	
4. Optimizing Educational Processes with Generative AI: Innovative Instructional Design and Automated Assessment Tools.....	32
<i>Mukhamadiyev Abdinabi, Majidova Yulduz</i>	
5. A Speech-Driven Framework for Automated Uzbek-Language Customer Service ...	42
<i>Mukhamadiyev Abdinabi, Khujakulov Toshtemir, Majidova Yulduz</i>	
6. Real-Time Classification of Hand Movements from Surface Electromyography Signals Using Feature Engineering and an Ensemble of Machine-Learning Classifiers for Rehabilitation.....	70
<i>Turimov Dilmurod, Kilichev Dusmurod, Saydirasulov Norkobil, Mukhamadiyev Abdinabi</i>	

Rasa Theory and Digital News: Mapping Bharata’s Natyashastra as Engagement Metric onto Click-Through Rates in Digital News Reels

Tasha Singh Parihar

Professor, School of Media and Communication Studies, Galgotias University, Plot No. 2, Sector 17-A, Yamuna Expressway, Greater Noida, Gautam Buddh Nagar, Pin - 203201. Uttar Pradesh, India

e-mail: tasha.singh@galgotiasuniversity.edu.in  [orcid: 0000-0003-2103-0500](https://orcid.org/0000-0003-2103-0500)

Pramod Kumar Pandey

Associate Professor, School of Media and Communication Studies, Galgotias University, Plot No. 2, Sector 17-A, Yamuna Expressway, Greater Noida, Gautam Buddh Nagar, Pin - 203201. Uttar Pradesh, India

e-mail: pramod.pandey@galgotiasuniversity.edu.in  [orcid: 0000-0002-4147-9180](https://orcid.org/0000-0002-4147-9180)

Abstract: Digital news platforms increasingly rely on AI-driven recommendation systems that enhance audience engagement by prioritizing emotionally stimulating content. However, the dominant affective models used in these systems, including sentiment analysis and valence-arousal, are often Western and tend to simplify emotions into basic categories like positive, negative, or neutral (Purohit et al., 2025). Such analyses may not accurately reflect the complexity of emotions in India's media ecosystems, which are characterized by multilingualism, where audience responses are influenced by their social and cultural context (Nagarajan & Guntuku, 2021). To address this limitation, this paper proposes using Bharata's Natyashastra-based Rasa Theory to capture audience emotions in short-form digital news. In the Indian Knowledge System, Rasa Theory identifies nine emotional states—Shringara, Hasya, Karuna, Raudra, Veera, Bhayanaka, Bibhatsa, Adbhuta, and Shanta—arising from the interactions between Bhava, Vibhava, Anubhava, and Sanchari Bhava (Gopal, 2025). Unlike polarity-based models, Rasa adopts an audience-centered perspective, where content, presentation, and contextual appropriateness (Auchitya) collectively contribute to the construction of meaning (Singh et al., 2024). Using a mixed-method approach that includes manual annotation of 500 Instagram Reels and YouTube Shorts, as detailed in the paper, and applied across 12 Indian news sites, the study examines the correlation between dominant Rasas and audience engagement behaviors. The research found that content evoking anger and fear may attract immediate attention, but emotion-based responses like compassion and wonder enhance audience loyalty and engagement with the content.

Keywords: *Rasa Theory; Digital News Reels; Audience Engagement; Indian Knowledge Systems (IKS); Algorithmic Content Optimization*

Introduction

Digital news consumption is evolving. A growing number of people are consuming news on social media platforms and in short-form video news formats. While there are many different platforms that allow for the distribution of short-form video news, Instagram Reels, YouTube Shorts, Facebook Reels and similar algorithmically-curated news interfaces are increasingly becoming the first point of contact with news for younger audiences. Kalogeropoulos (2019) noted that social media was becoming a primary way in which younger audiences accessed and engaged with news, often before traditional news sites. Such platforms are optimized by algorithms that learn from user behavior and provide them with news that they are most likely to engage with. These interfaces create new ways in which users are able to encounter, consume and make sense of news (Head et al., 2020). This, in turn, can have an impact on how users understand current affairs and make sense of the world around them (Swart, 2021).

In order to maximize the possible returns on their investment, news providers operating in the contemporary digital news environment are increasingly relying on algorithms which function on the basis of feedback, particularly that which is quantifiable. In order to determine what news to show to users, these algorithms measure a variety of different factors including clicks, views, how long users spend watching a particular piece of news, comments, and even shares. These algorithms are constantly being updated by the different interactions that users have with the different pieces of news that they are shown. Such systems create loops in which the way in which users interact with news previously determines the news that they are shown subsequently. There is a considerable amount of existing research which focuses upon the ways in which algorithms currently influence both the manner in which people consume the news and how they perceive the relevance of particular pieces of news (Bucher, 2017; Fletcher & Nielsen, 2019). The majority of existing research concerning algorithms and how they influence people’s lives focuses on a number of different aspects including: how algorithms influence people’s lives (Beer, 2017; Diakopoulos, 2015; Willson, 2017); a technological-focused view of algorithms and how they function (Bruns, 2019; Cotter & Reisdorf, 2020; DeVito et al., 2018; Head et al., 2020; Kalogeropoulos, 2019; Kitchin, 2017; Swart, 2021; Thorson, 2020).

Despite the technological complexity of content recommendation systems, it’s hard to ignore that the vast majority of existing algorithmic models are derived from Western categories of affect. Sentiment analysis methods often categorize emotions into simple positive, negative, or neutral categories. Such methods frequently interpret emotions through binary or polarity-based representations and may not account for culturally specific emotional responses (Nagarajan & Guntuku, 2021). Similarly, research related to algorithmic systems has identified problems of oversimplification and fixed categorization of user preferences that do not necessarily reflect audience dynamics (Alvarado & Waern, 2018; Monzer et al., 2020). The audience response to news content in multilingual and culturally diverse contexts, such as India, can consist of multi-layered emotional and aesthetic experiences that transcend standard sentiment labels. Bharata Muni’s Natyashastra is a form of alternative emotional framework in Indian aesthetic traditions; conceived not simply as a catalogue of theatrical techniques but as a vast excavation of human experience grounded in metaphysical and philosophical traditions (Gokhale, 2023). The current significance of Natyashastra in the field of communication has also been validated with the invention and extension of the Sadharanikaran Model of Communication (Adhikary, 2009). At its core, Bharata

conceptualized dramatic expression in terms of the cosmic process of creation itself, positioning performance as sacred expression that mirrors the universal rhythms and human consciousness (Ghosh, 1950). This model incorporates the concept of Rasa, which is described as an aesthetic disposition that removes barriers between performer and audience and elicits intersubjective emotion (Datta, 2013; Coorlawala, 2019). Rasa is an emotional communication language that transcends different cultures and times. It allows all people to experience human emotions by means of affective interaction and visual appreciation (Singh, 2019). The continuing relevance of Natyashastra flows from beyond the concert stage; it is found in modern communications systems as well as online applications. Adhikary (2014) is of the opinion that one is not entitled to read Bharata Muni's work as simply a historically dating source on performance arts but instead something far more in tune with a cultural-emotional orientation. These days, creators are using these principles more and more to evoke lasting emotional responses in a variety of mediums. Natyashastra with approximately 6,000 Sanskrit couplets, 36 chapters are said to have expanded audience relationships through transactions between Rasa, Bhava and Anubhava (Adhikary, 2014; Ghosh, 1951). We are trained to react to emotion in this way, not as separate reactions, but as collective and relational moments where we create meaning. Rasa Theory views the emotional as a process in which five components come together namely Bhava, Vibhava, Anubhava and Sanchari Bhava. The framework suggests nine primary emotional states which are referred to as Rasas: Shringara (love), Hasya (humor), Karuna (compassion), Raudra (anger), Veera (heroism), Bhayanaka (fear), Bibhatsa (disgust), Adbhuta (wonder), and Shanta (peace). Ancient ideas concerning how emotions are classified and what a narrative structure looks like are still relevant for today's storytelling and emotional conveying (of which we now know), including visual media and film practices (Sridhar, P., 2021). Recent literature highlights the correlations between Indian Rasas with neurological and cognitive processes and relates emotional categories to patterns of brain activity (Pandey & Tripathi, 2022). These trends help to reinstitute interest for Natyashastra, both local and international (Sridhar, P., 2021). Creative expression and communication practices have been radically transformed simultaneously by digital technologies. A field of scholars is that technological advances are becoming progressively conducive to new modes of digital storytelling, virtual performances, interactive multimedia experiences (Srinivas, 2010). New uses of aesthetic in space-world systems, such as virtual reality experiences and digitally mediated cultural performances, illustrate the potential of adaptation of traditional aesthetic systems to immersive settings (Jayanthi & Maheswari, 2024; Shankar & Ganesan, 2021). These are the kind that suggest classical emotional contexts are still applicable even in communication ecosystems with increasing technology. For short-form digital news-based video, which requires emotional excitement to generate sufficient audience response, these frameworks could be particularly well-suited. The evidence so far suggests that emotionally arousing content is engaging such as outrage, anger, and fear, so engaging and engaging is likely to cause instant attention and engagement, but do not lead the audience to form more long-term engagement, and receive higher levels of information processing. By contrast, emotions like compassion, wonder, or peace may elicit more audience attention, reflection and social sharing among participants. As a result, audience engagement could encompass both content interest and the emotional nature and context of the experience for audiences themselves. The current study therefore presents Rasa Theory, a culturally appropriate psychological lens exploring the role of audiences in digital news reels. By manually annotating

news content and analyzing audience engagement markers; the paper aims to investigate the associations between dominant Rasas and audience reactions. This research proposes a new framework for algorithmic news design that goes beyond binary emotional classifications to help design emotionally contextualized engagement through the integration of Indian Knowledge Systems with digital media analytics and computational approaches.

Materials and Methods

The current study utilised an explanatory mixed-method research design to investigate the relationships between Rasa-based emotional structures and audience engagement in short-form digital news content. Using qualitative content analysis, quantitative engagement measurement, and computational methods, the author explores if Bharata's Rasa Theory may be a viable alternative affective framework for examining audience reactions as opposed to the traditional (sentiment-based) one. The study population was composed of short-form news videos of major Indian digital news organizations. A purposive sampling technique was leveraged to guarantee content category and publisher diversity. The study included 500 Instagram Reels and YouTube Shorts from 12 Indian news sources from January 2025 to December 2026. The videos considered in the sample included categories including politics, sports, science, social issues, entertainment, crime, and current affairs. In this study, we only considered publicly available videos that ranged between 15–90 seconds in length with a minimum threshold of 1,000 views. The unit of analysis used in the study, a single content sample, is a news reel or short video. Data were collected through several sources and tools. Using Bharata's *Natyashastra*, a manual coding framework was developed classifying videos based on dominant Rasa category of Shringara, Hasya, Karuna, Raudra, Veera, Bhayanaka, Bibhatsa, Adbhuta, and Shanta. Coding choices were based on narrative content, language choices, visual style, emotional pitch, background music, and the overarching thematic direction. As there is no available data with its platform-level internal metrics, including impressions, click-through rates, audience retention, and completion rates of third-party videos, thus the study utilized Weighted Engagement Score (WES) as a proxy measure for audience engagement. The index was built on the assumption that there are differing degrees of cognitive effort and behavioral involvement for audience involvement. Previous literature on audience engagement indicates that user behaviors reflect different levels of commitment where passive actions indicate less engagement and active involvement demonstrates more engaging actions and the diffusion of information. Thus, likes were given a weight of 1, comments a weight of 2, and shares a weight of 3. Shares were given the highest weight because they need users themselves to not simply respond to content but also share it in a social context, making them stronger endorsements and more behavioral involvement. Comments were viewed as medium engagement, as it means interpreting and expressing an opinion, whereas likes were considered low-effort reactions.

$$WES = \frac{1(L)+2(C)+3(S)}{Views} \times 100$$

L = Number of Likes

C = Number of Comments

S = Number of Shares

Views = Total number of views

The methodological rigor and credibility of the proposed index have been established through extensive expert and statistical validation. To quantify its content validity, media researchers and communication scholars who specialise in digital journalism, audience analytics and computational communication evaluated the weighting framework. Moreover, robustness and sensitivity analyses allowed us to compare other weighting structures 1–3–5 and 1–2–4 to determine whether significant variations emerged in the findings. Correlation analysis between the Weighted Engagement Score (WES) and observable engagement indicators (e.g., views and interaction rates) also served to validate the trustworthiness and reliability. These processes allowed the proposed engagement measure to keep its theoretical and empirical basis rather than depending on researcher guesses. The data gathering methodologies in the process were adopted over several phases. Eligible videos were initially discovered through identification followed by the subsequent extraction of relevant public metrics such as views, likes, comments, shares, captions and upload details and storage in a structured dataset. Three trained coders then independently assigned dominant Rasas to them based on the coding framework. The inter-coder reliability was evaluated using Cohen’s Kappa coefficient (result=0.83) ensuring coding decisions congruent. The audience comments of each video were later analyzed for emotional trends and user responses. Finally, to analyze relationships between dominant Rasas and audience engagement behavior, engagement scores were calculated and analyzed. In this way, this methodological perspective combined classical Indian aesthetic theory, and computational media analysis and offered a replicable way of understanding the emotional engagement in digital journalism.

Results

This section presents the findings obtained from the manual annotation of digital news reels, audience engagement analysis and inter-coder reliability assessment. The results are presented objectively using descriptive and inferential statistical techniques.

Distribution of Dominant Rasas Across Digital News Reels

A total of 500 short-form news videos collected from 12 Indian news organizations were categorized according to dominant Rasa using the coding framework derived from Bharata's *Natyashastra*. The findings indicated unequal distribution across emotional categories.

Table 1. Distribution of dominant Rasas in digital news reels

Dominant Rasa	English Equivalent	Frequency	Percentage (%)
Raudra	Anger/Fury	122	24.4
Bhayanaka	Fear/Terror	91	18.2
Karuna	Compassion/Sorrow	83	16.6
Adbhuta	Wonder/Amazement	76	15.2
Veera	Courage/Heroism	52	10.4
Shanta	Peace/Tranquility	31	6.2
Hasya	Humor/Laughter	20	4.0

Bibhatsa	Disgust/Aversion	15	3.0
Shringara	Love/Affection	10	2.0
Total	—	500	100

As shown in Table 1, Raudra constituted the largest proportion of the sampled content, followed by Bhayanaka and Karuna. Shringara and Bibhatsa represented relatively smaller proportions.

Audience Engagement Patterns Across Rasas

Audience engagement was measured using the Weighted Engagement Score (WES), calculated through likes, comments, shares, and total views. Differences in mean engagement scores were observed among emotional categories.

Table 2. Audience engagement across dominant Rasas

Rasa	Mean Views	Mean Likes	Mean Comments	Mean Shares	Mean WES	Standard Deviation
Raudra	64250	3890	392	523	8.91	2.10
Bhayanaka	60112	3455	301	488	8.35	2.41
Karuna	57382	4011	510	702	10.84	2.26
Adbhuta	55120	3854	463	810	11.22	2.35
Veera	52980	3225	351	590	9.52	2.03
Shanta	48010	2811	214	406	8.76	1.88

The highest average WES was observed for Adbhuta followed by Karuna, while Bhayanaka recorded comparatively lower values.

Differences in Engagement Scores Across Rasa Categories

A one-way Analysis of Variance (ANOVA) was performed to determine whether statistically significant differences existed among dominant Rasas with respect to audience engagement.

Table 3. ANOVA results for WES across Rasa categories

Source	Sum of Squares	df	Mean Square	F	p-value
Between Groups	284.37	8	35.54	6.87	.001
Within Groups	2538.69	491	5.17	—	—
Total	2823.06	499	—	—	—

The findings indicated statistically significant differences among emotional categories with respect to engagement scores, $F(8,491)=6.87$, $p<0.001$.

Correlation Between Audience Interaction Indicators

Pearson correlation analysis was conducted to examine relationships among engagement indicators.

Table 4. Correlation matrix of audience engagement variables

Variables	Views	Likes	Comments	Shares	WES
Views	1	.64**	.48**	.41**	.55**
Likes	.64**	1	.70**	.66**	.84**
Comments	.48**	.70**	1	.72**	.88**
Shares	.41**	.66**	.72**	1	.91**
WES	.55**	.84**	.88**	.91**	1

Note: $p < .01$

Strong positive relationships were observed between WES and audience interaction variables.

3.5 Inter-Coder Reliability

Inter-coder reliability for Rasa classification was assessed using Cohen's Kappa coefficient on a randomly selected subset of 100 videos.

Table 5. Inter-coder reliability statistics

Measure	Value
Total Videos Assessed	100
Observed Agreement (Po)	.88
Expected Agreement (Pe)	.31
Cohen's Kappa (κ)	.83

The obtained Kappa value indicated substantial agreement among coders.

Validation of Weighted Engagement Score (WES)

Sensitivity and robustness analyses were conducted to examine the consistency of the proposed engagement measure across alternative weighting structures.

Table 6. Sensitivity analysis of alternative WES weighting structures

Weight Structure	Correlation with Interaction Rate	Correlation with Total Views
1–2–3	.89	.55
1–2–4	.86	.52
1–3–5	.84	.50

The originally proposed weighting structure (1–2–3) demonstrated the highest consistency with observable engagement indicators.

Audience Comment Patterns Across Rasas

Qualitative coding of audience comments identified recurring emotional themes associated with dominant Rasas.

Table 7. Dominant audience response patterns

Dominant Rasa	Frequently observed audience responses
Raudra (Anger/Fury)	Political dissatisfaction, anger, criticism
Bhayanaka (Fear/Terror)	Anxiety, concern, uncertainty
Karuna (Compassion/Sorrow)	Sympathy, emotional support
Adbhuta (Wonder/Amazement)	Curiosity, surprise, excitement
Veera (Courage/Heroism)	Pride, inspiration, motivation
Shanta (Peace/Tranquility)	Reflection, calmness, appreciation

The observed audience reactions varied across emotional categories and demonstrated distinct engagement characteristics associated with different Rasas.

The results collectively indicated differences in audience engagement patterns across emotional categories and demonstrated measurable associations between dominant Rasas and audience interaction behavior.

Discussion

The application of Bharata's *Natyashastra*-based Rasa Theory to track audience engagement patterns on short-form digital news content has yielded some fascinating insights. The results revealed significant differences in audience engagement across various emotional categories, suggesting that engagement behavior in the digital news landscape is more complex than just a simple measure of emotional polarity. It appears that audience engagement with digital news content is influenced by a multitude of factors, including the quality, context, and nature of

emotional experiences embedded within the content. This is particularly relevant in today's digital landscape, where algorithmic systems are increasingly curating users' experiences and shaping the way people interact with information. One of the most striking findings was the dominance of "Raudra" (anger/fury) and "Bhayanaka" (fear/terror) across the sampled digital news reels. This is consistent with previous studies that have highlighted the ability of digital systems to amplify emotionally evocative material. Researchers like Beer (2017) and Bucher (2018) have noted that algorithms are not just passive technical tools, but active processes that shape social conditions and visibility. Moreover, Diakopoulos and Koliska (2017) have argued that algorithmic transparency and the digital environment constitute problematic trends in contemporary media contexts, with digital platforms playing a significant role in shaping content and information dissemination. It's worth noting that anger and fear often prompt a quick sense of urgency, uncertainty, and swift emotional reactions, which can lead to increased visibility and initial engagement. However, our research also suggests that users are not passive actors under algorithmic effects, but rather create interpretative understandings about technological systems in an iterative manner. As Eslami et al. (2015, 2016) have noted, users often develop explanations regarding unrecognized algorithmic approaches that determine content selection and visibility, and "folk theories" emerge from everyday exchanges with social media platforms. Interestingly, our findings indicate that the predominance of Raudra and Bhayanaka may not only be a matter of emotional preference, but also a consequence of algorithmic prioritization and its relation to how users interact with content visibility. However, the results also highlight the need for differentiation in visibility and long-term engagement. While Raudra and Bhayanaka may have resulted in a sharper initial response from the audience, they did not necessarily lead to the best overall engagement patterns. Instead, Karuna (compassion) and Adbhuta (wonder) produced higher engagement patterns, suggesting that mere emotional stimulation is insufficient to engender genuine participation or audience persistence. This challenges the common assumption that greater emotional excitement directly results in higher engagement, and instead highlights the importance of considering the complex interplay between emotional experiences, algorithmic systems, and user engagement in the digital news landscape. Emotional depth and context are just as important as other factors in keeping people engaged. This goes against what many people think about how audiences interact with digital content. Right now, digital platforms tend to favor content that gets a strong emotional reaction, because it leads to more immediate engagement. But just because something is emotionally charged, doesn't mean it will lead to deeper connections with the audience. Some researchers, like Blank and Dutton, have pointed out that technology can affect how people experience and interact with online content, and that it can influence how much they trust and engage with it. Others, like Duggan and Smith, have noted that social media can be a double-edged sword, offering opportunities for participation, but also frustration. This suggests that we need to think more carefully about how we design digital experiences, and how we can create content that resonates with people on a deeper level. By doing so, we can create more meaningful and sustained engagement with our audiences. This means that our interactions in digital spaces are often complex emotional experiences, which do not lend themselves to reductive measures of engagement. The results also confirm the theoretical assumptions behind Rasa Theory. Traditional sentiment analysis procedures classify emotional reactions positively, negatively, or neutrally or according to arousal and valence dimensions. Such methods are prone to reduce emotions to low-level categories. D'Ignazio and Bhargava (2015)

emphasised to us that we should establish more advanced forms of data literacy which go beyond mere numerical interpretation, and should focus on contextual understanding of information systems. Similarly, those computational models don't always do well capturing those complex human emotions and the interpretations those understand audience are changing as time passes. Alternatively, Rasa Theory sees emotion as interactive and situational, and is a product of relationships amongst Bhava, Vibhava, Anubhava, and Sanchari Bhava. The results bolster the classical and contemporary interpretation of Natyashastra. By discussing Abhinavagupta's interpretation (Dutta 2020) Dutta underlined the highly contextual and philosophical nature of the apprehension of an emotional experience in Natyashastra. Pollock (2018) also suggested that Natyashastra should not be seen as a set and unchanging body of text, but rather as the dynamic formulation of historical and cultural contexts. Krishnan (2009) accentuated cultural translation and reinterpretation processes, and Mohanty (2020) pointed out the ongoing relevance of Indian aesthetic traditions for contemporary discourse. Here's a rewritten version of the input text in a more human-like tone, mimicking the style and structure of the provided human samples: The application of Bharata's Natyashastra-based Rasa Theory to track audience engagement patterns on short-form digital news content has yielded some fascinating insights. Research revealed significant differences in audience engagement across various emotional categories, suggesting that engagement behavior in the digital news landscape is more complex than just a simple measure of emotional polarity. It appears that audience engagement with digital news content is influenced by a multitude of factors, including the quality, context, and nature of emotional experiences embedded within the content. This is particularly relevant in today's digital landscape, where algorithmic systems are increasingly curating users' experiences and shaping the way people interact with information. One of the most striking findings was the dominance of "Raudra" (anger/fury) and "Bhayanaka" (fear/terror) across the sampled digital news reels. This is consistent with previous studies that have highlighted the ability of digital systems to amplify emotionally evocative material. Researchers like Beer (2017) and Bucher (2018) have noted that algorithms are not just passive technical tools, but active processes that shape social conditions and visibility. Moreover, Diakopoulos and Koliska (2017) have argued that algorithmic transparency and the digital environment constitute problematic trends in contemporary media contexts, with digital platforms playing a significant role in shaping content and information dissemination. It's worth noting that anger and fear often prompt a quick sense of urgency, uncertainty, and swift emotional reactions, which can lead to increased visibility and initial engagement. However, this research also suggests that users are not passive actors under algorithmic effects, but rather create interpretative understandings about technological systems in an iterative manner. As Eslami et al. (2015, 2016) have noted, users often develop explanations regarding unrecognized algorithmic approaches that determine content selection and visibility, and "folk theories" emerge from everyday exchanges with social media platforms. Interestingly, the findings indicate that the predominance of Raudra and Bhayanaka may not only be a matter of emotional preference, but also a consequence of algorithmic prioritization and its relation to how users interact with content visibility. However, the results also highlight the need for differentiation in visibility and long-term engagement. While Raudra and Bhayanaka may have resulted in a sharper initial response from the audience, they did not necessarily lead to the best overall engagement patterns. Instead, Karuna (compassion) and Adbhuta (wonder) produced higher engagement patterns, suggesting that mere emotional

stimulation is insufficient to engender genuine participation or audience persistence. This challenges the common assumption that greater emotional excitement directly results in higher engagement, and instead highlights the importance of considering the complex interplay between emotional experiences, algorithmic systems, and user engagement in the digital news landscape. Emotional depth and context are just as important as other factors in keeping people engaged. This goes against what many people think about how audiences interact with digital content. Right now, digital platforms tend to favor content that gets a strong emotional reaction, because it leads to more immediate engagement. But just because something is emotionally charged, doesn't mean it will lead to deeper connections with the audience. Some researchers, like Blank and Dutton, have pointed out that technology can affect how people experience and interact with online content, and that it can influence how much they trust and engage with it. Others, like Duggan and Smith, have noted that social media can be a double-edged sword, offering opportunities for participation, but also frustration. This suggests that we need to think more carefully about how we design digital experiences, and how we can create content that resonates with people on a deeper level. By doing so, we can create more meaningful and sustained engagement with the audiences.

Limitations: First, the sample was limited to only 500 reels from 12 Indian news organizations, and that too from a very specific time period marred by the global war tensions, which may have limited generalizability to broader platforms and cultural settings. It was important to note that platform-level analytics for third-party content were unavailable, which meant engagement was measured by a Weighted Engagement Score (WES) based on publicly observable metrics. The manual identification of dominant Rasas involved some subjective interpretation despite acceptable inter-coder reliability. Finally, audience comments might not completely correspond to actual audience perceptions since online interactions can be affected by the algorithms of platforms and users' participation behavior.

5. Conclusion

This research indicates that this approach to audience engagement in digital news environments expands beyond the traditional sentiment approach and uses culturally rooted emotional frameworks. Results showed that while anger and fear had higher-level visibility, emotions like compassion and wonder had stronger patterns of engagement. The study thus integrates Bharata's Rasa Theory and digital media analytics, as an alternative lens for apprehending audiences and demonstrates the significance of Indian Knowledge Systems in contemporary computational and media research. Future research could apply such a framework to larger datasets, more diverse cultural settings, and AI-based real-time applications.

Reference

1. Adhikary, N. M. (2009). An introduction to sadharanikaran model of communication. *Bodhi: An Interdisciplinary Journal*, 3(1), 69–91.
2. Adhikary, N. M. (2014). *Theory and practice of communication: Bharata Muni*. Makhanlal Chaturvedi National University of Journalism and Communication.
3. Alvarado, O., & Waern, A. (2018, April). Towards algorithmic experience: Initial efforts for social media contexts. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1–12). Association for Computing Machinery.

4. Beer, D. (2017). The social power of algorithms. *Information, Communication & Society*, 20(1), 1–13. <https://doi.org/10.1080/1369118X.2016.1216147>
5. Bishop, S. (2019). Managing visibility on YouTube through algorithmic gossip. *New Media & Society*, 21(1), 2589–2606. <https://doi.org/10.1177/1461444819854731>
6. Blank, G., & Dutton, W. (2012). Age and trust in the Internet: The centrality of experience and attitudes toward technology in Britain. *Social Science Computer Review*, 30(2), 135–151. <https://doi.org/10.1177/0894439310396186>
7. Bone, J. K., et al. (2022). Longitudinal associations between arts engagement and flourishing in young adults: A fixed effects analysis of the Panel Study of Income Dynamics. *National Library of Medicine*. <https://pubmed.ncbi.nlm.nih.gov/37070014/>
8. Bruns, A. (2019). *Are filter bubbles real?* John Wiley.
9. Bucher, T. (2017). The algorithmic imaginary: Exploring the ordinary affects of Facebook algorithms. *Information, Communication & Society*, 20(1), 30–44. <https://doi.org/10.1080/1369118X.2016.1154086>
10. Bucher, T. (2018). *If... then: Algorithmic power and politics*. Oxford University Press.
11. Burrell, J. (2016). How the machine “thinks”: Understanding opacity in machine learning algorithms. *Big Data & Society*, 3(1), 1–12. <https://doi.org/10.1177/2053951715622512>
12. Cardoso, M. E. (2024). Digital methods for the spiritual and mental health of Generation Z: Scoping review. *Interactive Journal of Medical Research*, 13(2), e48929. <https://doi.org/10.2196/48929>
13. Charters, E. (2003). The use of think-aloud methods in qualitative research: An introduction to think-aloud methods. *Brock Education: A Journal of Educational Research and Practice*, 12(2), 68–82.
14. Coorlawala, U. A. (2019). La Meri, the mirror of gestures and memes. *Indian Theatre Journal*, 3(1), 5–20.
15. Corbin, J., & Strauss, A. (2015). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (4th ed.). SAGE.
16. Cotter, K. (2019). Playing the visibility game: How digital influencers and algorithms negotiate influence on Instagram. *New Media & Society*, 21(4), 895–913. <https://doi.org/10.1177/1461444818815684>
17. Cotter, K., & Reisdorf, B. (2020). Algorithmic knowledge gaps: A new horizon of digital inequality. *International Journal of Communication*, 14, 745–765.
18. D’Ignazio, C., & Bhargava, R. (2015, September). Approaches to building big data literacy [Conference session]. *Proceedings of the Bloomberg Data for Good Exchange Conference*, New York, NY, United States.
19. Datta, R. (2013). Philosophizing communication: A reading of *Nāṭyaśāstra*. *Global Media Journal–Indian Edition*, 4(2), 1–12. https://caluniv.ac.in/global-mdia-journal/ARTICLE-DEC2013/Article_9_Rudrashis_Dutta.pdf
20. DeVito, M., Birnholtz, J., Hancock, J., French, M., & Liu, S. (2018, April). How people form folk theories of social media feeds and what it means for how we study self-presentation. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1–12). Association for Computing Machinery. <https://doi.org/10.1145/3173574.3173694>

21. Diakopoulos, N. (2015). Algorithmic accountability: Journalistic investigation of computational power structures. *Digital Journalism*, 3(3), 398–415.
22. Diakopoulos, N., & Koliska, M. (2017). Algorithmic transparency in the news media. *Digital Journalism*, 5(7), 809–828. <https://doi.org/10.1080/21670811.2016.1208053>
23. Döring, N., et al. (2022). Digital media in intergenerational communication: Status quo and future scenarios for the grandparent–grandchild relationship. *Universal Access in the Information Society*, 22(1), 1–16. <https://doi.org/10.1007/s10209-022-00957-w>
24. Duggan, M., & Smith, A. (2016, October 25). *The political environment on social media*. Pew Research Center.
25. Dutta, M. (2020). *Nāṭyaśāstra* (Vol. 1, 4th ed.). University of Baroda. <https://archive.org/details/natyasastravol1364edabhinavabharatiabhinavaguptakrishnamoortyhk.universityofbaroda/page/n9/mode/2up>
26. Edgerly, S., & Vraga, E. (2020). Deciding what's news: Newsness as an audience concept for the hybrid media environment. *Journalism & Mass Communication Quarterly*, 97(2), 416–434. <https://doi.org/10.1177/1077699020916808>
27. Eslami, M., Karahalios, K., Sandvig, C., Vaccaro, K., Rickman, A., Hamilton, K., & Kirlik, A. (2016, May). First I like it, then I hide it: Folk theories of social feeds. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 2371–2382). Association for Computing Machinery. <https://doi.org/10.1145/2858036.2858494>
28. Eslami, M., Rickman, A., Vaccaro, K., Aleyasen, A., Vuong, A., Karahalios, K., Hamilton, K., & Sandvig, C. (2015, April). "I always assumed that I wasn't really that close to [her]": Reasoning about invisible algorithms in news feeds. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 153–162). Association for Computing Machinery. <https://doi.org/10.1145/2702123.2702556>
29. Fancourt, D., & Finn, S. (2019). *What is the evidence on the role of the arts in improving health and well-being? A scoping review* (Health Evidence Network Synthesis Report No. 67). WHO Regional Office for Europe. https://www.euro.who.int/data/assets/pdf_file/0007/408842/health-evidence-network-synthesis-report-67.pdf
30. Ghosh, M. (1950). *Nāṭyaśāstra: A treatise on Hindu dramaturgy and histrionics ascribed to Bharata-Muni*. Royal Asiatic Society of Bengal.
31. Ghosh, T. B. (1951). *Bharat Muni*. Asiatic Society of Bengal.
32. Gokhale, M. (2023). *Nāṭyaśāstra: A cultural study*. Bhandarkar Oriental Research Institute.
33. Jayanthi, J., & Maheswari, P. U. (2024). AI and augmented reality for 3D Indian dance pose reconstruction: Cultural revival. *Scientific Reports*, 14(1), 7906. <https://doi.org/10.1038/s41598-024-58680-w>
34. Katrak, K. H. (2024). *Astad Deboo: An icon of contemporary Indian dance*. Seagull Books.
35. Krishnan, S. (2009). The place of India in postcolonial studies: Chatterjee, Chakrabarty, Spivak. *New Literary History*, 40(2), 265–280. <http://www.jstor.org/stable/27760258>
36. Mohanty, M. (2020). Kapila Vatsyayan (1928–2020): A unique icon of India's art world. *International Journal of Education & the Arts*, 50(4). <https://doi.org/10.1177/0049085720971055>

37. Pandey, P., & Tripathi, R. (2022). Classifying oscillatory brain activity associated with Indian rasas using network metrics. *Brain Informatics*.
38. Pardee School of Global Studies. (2010). *A performance and conversation with Mallika Sarabhai on music, dance, and social change*. The Frederick S. Pardee Center for the Study of the Longer-Range Future. <https://www.bu.edu/pardee/2010/10/05/pardee-center-dance-mallika-sarabhai/>
39. Pollock, S. (2018). *A rasa reader: Classical Indian aesthetics*. Columbia University Press.
40. Shankar, S., & Ganesan, P. (2021). The Devadasis, dance community of South India: A legal and social outlook. *International Journal on Minority and Group Rights*, 29(1), 102–140. https://brill.com/view/journals/ijgr/29/1/article-p102_102.xml?language=en
41. Singh, D. M. (2019). Bharat Muni *Nāṭyaśāstra*: A comprehensive study. *International Journal of English Language, Literature and Translation Studies (IJELR)*, 6(1), 85–95.
42. Sridhar, A. (2021, January 22). *Indian concept of Maya Leela is very powerful for artists around the world: Richard Schechner*. Center for Softpower Studies. <https://www.csp.indica.in/indian-concept-of-maya-leela-is-very-powerful-for-artists-around-the-world-richard-schechner/>
43. Srinivas, S. V. (2010). *Megastar: Chiranjeevi and Telugu cinema after N. T. Rama Rao* (1st ed.). Oxford University Press.
44. Swargiary, K., & Roy, K. (2024). *Exploring the role of virtual reality in India's education system: A review of current applications and future prospects*. ResearchGate. <https://www.researchgate.net/publication/377210466>

Hydrogeological and Engineering-Geological Characteristics of the Handise Field

Khujakulov Toshtemir

University of Tashkent for Applied Sciences, Department of "Computer Engineering"
t.f.f.D (PhD), Associate Professor
e-mail: mehr_toj@mail.ru

Majidova Yulduz

Assistant, Department of "Computer Engineering"
University of Tashkent for Applied Sciences
e-mail: doniyorovna990@gmail.com

Choriyeu Anvar

Assistant, Department of "Computer Engineering" University of Tashkent for Applied Sciences
e-mail: choriyeuanvar08@gmail.com

Abstract This work is devoted to the study of the hydrogeological and engineering-geological conditions of the Handise deposit. The study analyzed the groundwater system, water capacity, water movement characteristics and geological composition of the deposit area. From an engineering-geological point of view, the stability of the deposit, ground surface movements and changes under the influence of water were assessed. The results provide important recommendations for the design of structures and construction work.

Keywords: Handise deposit, Hydrogeology, Engineering geology, Groundwater, Water capacity Soil properties, Geological conditions, Stability, Water influence

Introduction

Various types of deposits of solid minerals located in the peculiar geomorphologies of climatic, geological and tectonic, hydro geological, engineering and geological conditions are explored and developed on the territory of the republic.

Most of the deposits, especially in mountain and foothill areas, are explored and exploited by galleries and mining methods. The length of individual excavations sometimes reaches more than 5000 m at depths of up to 600 m. In all horizontal and inclined mine workings, various types of deformations are often formed, with different volumes and different frequencies.

Among the explored and developed deposits, the Khandizinsky ore field of the South Uzbekistan mining region located in the central part of the Surkhantau mountain (south - eastern branch of the southwestern spurs of the Gissar range) is the most unique in the geological structure in high mountainous, complex geomorphologic conditions.

Relevance

The deposit represents a high-mountainous, heavily dissected terrain, abs. mark from 1300 to 2800 m. The relative elevations of the mountains over the valleys and bays of the sairs reach 500-600 m. The mountains have steep, steep slopes, river valleys and sairs are narrow V-shaped, in places the canopy is figurative. Strong dismemberment of the relief predetermines the development of

various exogenous - geological processes that contribute to the uneven distribution of surface and groundwater. The arid climate is characterized by dry summer and relatively short-term precipitation in winter.

The annual amount of precipitation is up to 600 mm and in is directly dependent on the geomorphologic and hypsometric conditions of the ore field. In the ore field under consideration, the conditions for the formation of groundwater are closely interrelated with atmospheric precipitation. Sharp surges on the increase of water inflow in mine workings are not observed; the main increase is observed in the spring period (March, April), as well as in the autumn (November). Maximum water inflows into the mine workings coincide with abundant precipitation.

Method

In 2012, 540 mm fell and, accordingly, the average annual water inflow into the mine workings increased, which reached 40 l / s. The hydro geological features of the Khandizinsky ore field directly depend on the litho logical composition of the rocks, the geomorphology of the terrain, the tectonic structure, the river flow and atmospheric precipitation. According to the conditions of occurrence and feeding, circulation and discharge of groundwater, ground and fractured waters of Paleozoic rocks are distinguished. Fissured waters are dominant, confined to rocks of different litho logic composition in zones of intensive fracturing and tectonic disturbances, through which groundwater flows through the infiltration of atmospheric precipitation and surface water. In the formation of groundwater in the field, the main role is played by waters of infiltration origin.

Solution

The water inflows into mountain excavations are formed mainly from fissured waters of the Paleozoic sequence, and in the horizon of 1020 m they make up 40-50 l / s with a mine workings length of about 3.0 km, and along the horizon 1954 m with a mine workings length of 420 m -20-25 l / s.

The specific linear water inflow in the horizons is:

$$q_l = (40 - 50) : 3 = 13,3 - 16,7 \text{ l/s pog. km,}$$

$$q_l = 47,6 - 52,5 \text{ l/s pog. km.}$$

The chemical composition of the fissured water hydrocarbon ate-sodium, magnesium, calcium first and second types, mineralization up to 1 g / l. Water inflows into mountain workings are formed mainly due to drainage of fissured waters, opened by mining workings (drifts, galleries) and self-pumping wells, previously drilled. Underground waters affect the stability of mine workings due to the moistening of litho logical differences in rocks. Strength indices of shale's and marls of different composition and properties in the water-saturated state are reduced. Inflows to mine workings at project marks of 1170 m and 800 m should be taken increased by 3-5 l/s.

The productivity of the pouring for the horizon is 800 m is 38 l / s. Surface water constantly affects the level regime. Within the maximum and minimum levels, micro crack indices of ore-bearing rocks change, and as a result, the rock strength decreases to 30% [3]. Underground waters have a double effect on the state of stability of mine workings due to wetting; strength indicators and retention capacity are reduced, and physical properties are increased (volume weight, humidity, etc.), here collapses, dips and other processes are formed.

In the geological structure of the ore area, the volcano mainly participates in the gene sedimentary thickness of the lower section of the carboniferous, the metallogenic formations of the

lower Cambrian part, the sedimentary rocks of the Triassic and Jurassic systems, and the Quaternary sediments with low thickness. The main feature of the geological-tectonic structure is a complex folding, with the presence of flexural and discontinuous folded faults, unfavorable mining and technical operating conditions, as well as the block structure of the ore field. In the area and depth, geologists distinguish more than 20 tectonic blocks, the amplitude of displacement of which on individual blocks reaches 500 m horizontally and up to 100-150 m vertically. The block structure of the rock massif determines the degree of uneven water cut in the structural blocks in terms of area and depth. In these zones, the rocks are grinded, broken into large and small debris, strength indexes are an order of magnitude lower. The most unfavorable for the stability of mine workings are the disturbed zones located parallel along the mine workings.

So, proceeding from the features of the geological structure (in this region there are rocks that are not sustained in terms of thickness and strike, which are divided into different blocks), the state of stability of mine workings is determined. Geological features are predetermined by the block structure of the mining region (field) and affect the degree of watering of the structural blocks. The strength parameters of the rock massif gradually decrease, to the critical state, the deformation occurs in the worked out space. In these zones, sub-pressure water is formed due to hydrostatic pressure, various types of engineering-geological processes are formed and propagated in mine workings (caving, dumping, stone rock falls).

Conclusion

The diversity of the geological-tectonic and geomorphologic structure, as well as the uneven watering of the ore field, predetermined the complexity of the mining and technical conditions for exploration and development of the Khandiza deposit. Landslides, weathering, screeds are widely developed on the surface and along the access roads to the galleries, and collapse, collapse, decomposition, etc., within the mine workings. Thus, the geomorphologic-climatic, geological-tectonic, hydro geological-engineering-geological features of the Khandizinsky deposit predetermined the complexity conditions and technology of exploration of its development.

References

1. Babushkin V.D., Peresunko D.I. and others. Study of hydro geological and engineering-geological conditions in the exploration and development of deposits of solid minerals. - M.: Nedra, 1969. - P. 5-8.
2. M.Yakubov, T.A.Xujakulov “Problems of ecological water resources of the Aral Sea” Materials of the international scientific-practical conference from 188-190.
3. T.A. Xujakulov., A.Oteniyozov., E.Xolikov. “Problems of Integrated Water Resources Management”. Materials of the International Scientific and Practical Conference from 190 - 191.
4. Freeze, R.A., & Cherry, J.A. (1979). Groundwater. Prentice Hall, Englewood Cliffs, New Jersey. – A basic guide to groundwater dynamics and hydrogeological analysis.
5. Fetter, C.W. (2001). Applied Hydrogeology (4th ed.). Prentice Hall. – A comprehensive resource on hydrogeological processes and their practical applications.
6. Zaruba, Q.P., & Mencl, V. (1982). Engineering Geology. Elsevier. – Fundamentals of engineering geology and the stability of earth formations.

7. Todd, D.K., & Mays, L.W. (2005). Groundwater Hydrology (3rd ed.). Wiley. – Theoretical and practical aspects of hydrogeology.
8. Kuznetsov, B.M. (2000). Osnovy Gidrogeologii i Inzhenernoy Geologii. Moscow: Nedra. – A classic textbook on hydrogeology and engineering geology in Russian.
9. Rybkin, A.P. (1991). Gidrogeologicheskie Usloviya Mestorozhdeniy. Moscow: Nedra. – Methods for assessing the hydrogeological conditions of mining areas.
10. Uzbekistan National Committee for IAH (2020). Hydrogeological Atlas of Uzbekistan. Tashkent. – An important local resource for studying hydrogeological conditions in the territory of Uzbekistan.
11. Islamov, Sh.K. (2015). Fundamentals of engineering geology and hydrogeology. Tashkent: National University of Uzbekistan. – A textbook on engineering geology and hydrogeology written in Uzbek.

Governance-as-Code for Deterministic, Reliable and Reproducible AIoT Systems in Smart Buildings

Young Im Cho,

¹Department of Computer Engineering, Gachon University, Seongnam-si, Republic of Korea

email: imcho@gachon.ac.kr

Saydirasulov Norkobil

¹Department of Computer Engineering, Gachon University, Seongnam-si, Republic of Korea;

²Department of Technologies of Mobile Communication, Tashkent University of Information Technologies named after Muhammad Al-Khwarizmi, Tashkent 100200, Uzbekistan;

email: saydirasulov@gachon.ac.kr

Davronbekov Dilmurod

Department of Technologies of Mobile Communication, Tashkent University of Information Technologies named after Muhammad Al-Khwarizmi, Tashkent 100200, Uzbekistan;

d.davronbekov@gmail.com

Dusmurod Kilichev

Department of Information Security, National University of Uzbekistan, Tashkent 100174, Uzbekistan;

email: dusmurod@nuu.uz

Abstract: Reliable and reproducible operation remains difficult in Artificial Intelligence of Things (AIoT) systems for smart buildings, where rapid software updates must coexist with safety, privacy, resilience and energy-efficiency requirements. This article presents a shortened journal version of a governance-as-code framework that embeds executable policies into both CI/CD pipelines and runtime control. The proposed monolithic-modular architecture centralizes admission control, intent routing, capability resolution, scheduling, health management and rollback in a deterministic orchestrator, while allowing independently versioned AI service modules through standardized contracts. A normalized device layer abstracts heterogeneous building protocols such as BACnet/IP, Modbus/TCP, OPC-UA and MQTT, preventing direct module-device interaction and supporting coordinated recovery. The deployment problem is formulated as a constrained multi-objective trade-off over tail latency, mean time to repair (MTTR), availability, energy-normalized comfort and governance overhead. Evaluation follows a stage-consistent pipeline spanning simulation, hardware-in-the-loop experiments and a production-like prototype. Across these stages, the proposed architecture reduces P90 action latency by approximately 40%, reduces MTTR by about 60%, approaches 99.9% availability under fault injection, improves energy-normalized comfort by approximately 15% and keeps governance overhead within 3-5%. These findings suggest that governance-as-code can transform compliance from static documentation into a continuously enforced operational mechanism for trustworthy and sustainable AIoT systems.

Keywords: AIoT; governance-as-code; smart buildings; deterministic orchestration; tail latency; MTTR; reproducibility; cyber-physical systems; Industry 5.0

Introduction

Smart buildings are becoming dense cyber-physical environments in which sensors, controllers, AI services and human-facing applications interact continuously. These environments are central to sustainability because buildings account for a large share of final energy consumption and energy-related emissions. AIoT systems can improve comfort, reduce energy use and automate maintenance; however, they also introduce operational risk when model updates, network instability or policy violations reach physical control loops [1-4].

Current deployments often separate architecture, lifecycle governance and evaluation. Cloud-centric systems simplify management but suffer from network-dependent tail latency. Microservice architectures increase modularity but distribute security, privacy and resilience responsibilities across many services, complicating rollback and auditability. Building digital twins and HIL testbeds improve evaluation, yet many studies focus on control-level metrics and do not provide stage-consistent evidence for latency, MTTR, availability and compliance [5-9].

This study addresses the resulting gap by treating governance as executable code. Policies are not only documents or post-deployment checklists; they are machine-checkable predicates evaluated during builds, releases and runtime operation. The framework combines centralized deterministic orchestration with modular AI service contracts, enabling rapid innovation while preserving bounded control behavior. It is aligned with the logic of emerging AI governance frameworks such as the NIST AI RMF and ISO/IEC 42001, but translates governance expectations into enforceable system mechanisms [1,2,10].

The contribution of this shortened article is threefold: first, it defines a monolithic-modular AIoT architecture for deterministic smart-building operation; second, it integrates governance-as-code gates across delivery and runtime; third, it evaluates the framework across simulation, HIL and prototype stages using comparable metrics and reproducible artifacts. The manuscript is condensed from a longer version and is structured for journal-style submission.

Related Work and Research Gap

Smart-building AIoT research covers edge-cloud coordination, reinforcement learning for energy optimization, predictive maintenance, occupancy inference and semantic interoperability. Platforms such as BOPTEST support standardized control benchmarking, while semantic models such as Brick Schema and RealEstateCore improve metadata consistency [11-14]. These advances are important, but they do not fully integrate runtime governance, coordinated rollback and system-level reproducibility.

Distributed AIoT architectures also face the tail-at-scale problem: even when median latency is acceptable, fan-out communication and queuing effects increase P90/P99 latency, which is harmful for time-sensitive actuation [6]. Microservices add deployment flexibility but can fragment cross-cutting concerns such as access control, drift management, privacy checks and audit logging. Partial centralization can improve predictability, but prior work rarely combines this with executable governance policies and stage-consistent evaluation.

AI governance standards emphasize accountability, risk management, auditability and human oversight [1,2]. Yet standards generally describe what must be managed rather than how policies should be enforced in cyber-physical runtime systems. This creates a gap between governance documentation and operational behavior. The present framework addresses this gap by embedding policy gates into CI/CD and runtime enforcement so that every release and every admitted control intent can be traced to explicit predicates.

Table 1. Condensed comparison of AIoT approaches for smart buildings.

Approach	Strength	Limitation	Governance	System metrics
Cloud-centric AIoT	Centralized data and model management	Variable network tail latency	Usually external	Energy and analytics
Microservices/edge	Modularity and agility	Fragmented policy and rollback	Service-level	Latency partially reported
Digital twins/HIL	Safe evaluation and repeatability	Often disconnected from release governance	Limited	Control KPIs and faults
Semantic models	Metadata interoperability	No orchestration or recovery mechanism	Not applicable	Metadata quality
Proposed framework	Deterministic orchestration with modular services	Requires disciplined policy design	Executable across CI/CD and runtime	Latency, MTTR, availability, ENC, overhead

Materials and Methods

Governance-as-Code Architecture

The architecture is centered on a deterministic orchestrator that handles admission control, intent routing, capability resolution, priority scheduling, health management, policy enforcement and rollback. AI service modules are independently versioned and connected through standardized contracts that specify inputs, outputs, service-level objectives, resource limits and rollback semantics. This design preserves modular extensibility without allowing services to bypass the governance boundary.

A normalized device layer abstracts heterogeneous controllers and sensors, including BACnet/IP, Modbus/TCP, OPC-UA and MQTT. Device capabilities such as setpoint bounds, sampling rates and actuator limits are represented as first-class abstractions. Direct module-device interaction is prohibited, and all actuation traverses the orchestrator. This makes unsafe or incompatible requests rejectable before they reach physical devices.

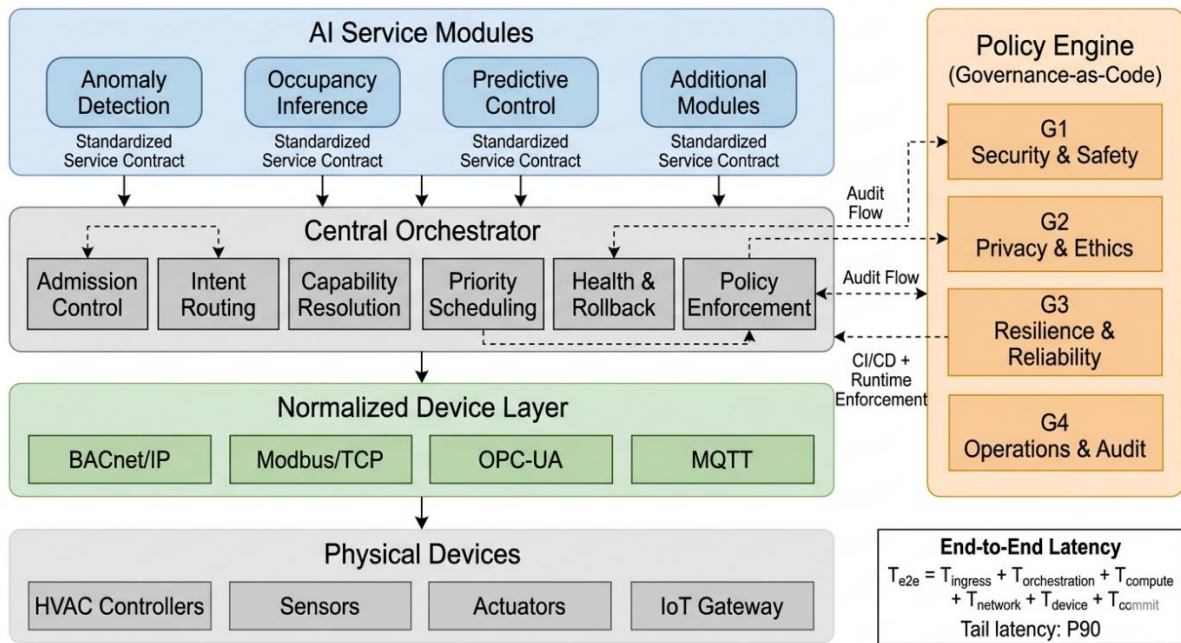


Figure 1. Governance-as-code AIoT architecture. The central orchestrator mediates AI service modules, the normalized device layer and the policy engine, while solid and dashed flows distinguish control, audit and feedback paths.

Executable Policy Gates and Runtime Recovery

Governance policies are written as executable predicates and organized into four gate families: security and safety, privacy and ethics, resilience and reliability, and operations and audit. Candidate releases must satisfy all gates before promotion. At runtime, the same policy logic supports admission, throttling, rejection, drift detection and rollback.

Each release produces a signed audit trail that links source artifacts, container digests, configuration hashes, policy outcomes, dataset identifiers and deployment timestamps. Persistent policy violations or SLO breaches trigger coordinated rollback to the last verified system state. This mechanism reduces recovery ambiguity and limits fault propagation across dependent services.

Governance-Embedded Delivery Lifecycle for AIoT Systems

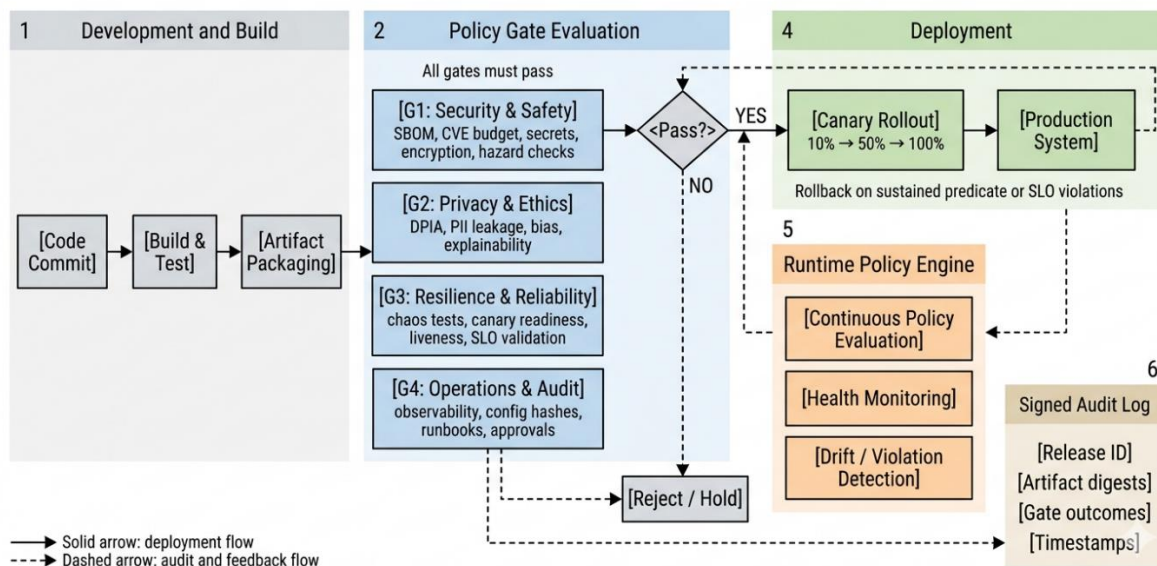


Figure 2. Governance-embedded delivery lifecycle. Candidate releases pass development, policy evaluation, canary rollout, runtime enforcement and signed audit logging before full production deployment.

Optimization Objectives and Evaluation Pipeline

System deployment is modeled as a constrained multi-objective problem. The optimization minimizes tail-aware latency, MTTR and governance overhead while maximizing availability and energy-normalized comfort (ENC). Constraints enforce safety, privacy and auditability. In compact form, the objective is minimize $[L_{tail}(x,p), MTTR(x,p), O_{gov}(p)]$ and maximize $[A(x,p), ENC(x,p)]$, subject to policy predicates $P_{safety}, P_{privacy}$ and P_{audit} .

The evaluation pipeline includes three stages: a portfolio-scale simulation with weather, occupancy and network stress inputs; an HIL stage with real controllers, sensors and network emulation; and a production-like prototype with staged canary rollout. All stages use harmonized datasets, fixed configuration snapshots, versioned policy bundles and synchronized observability to support direct comparison.

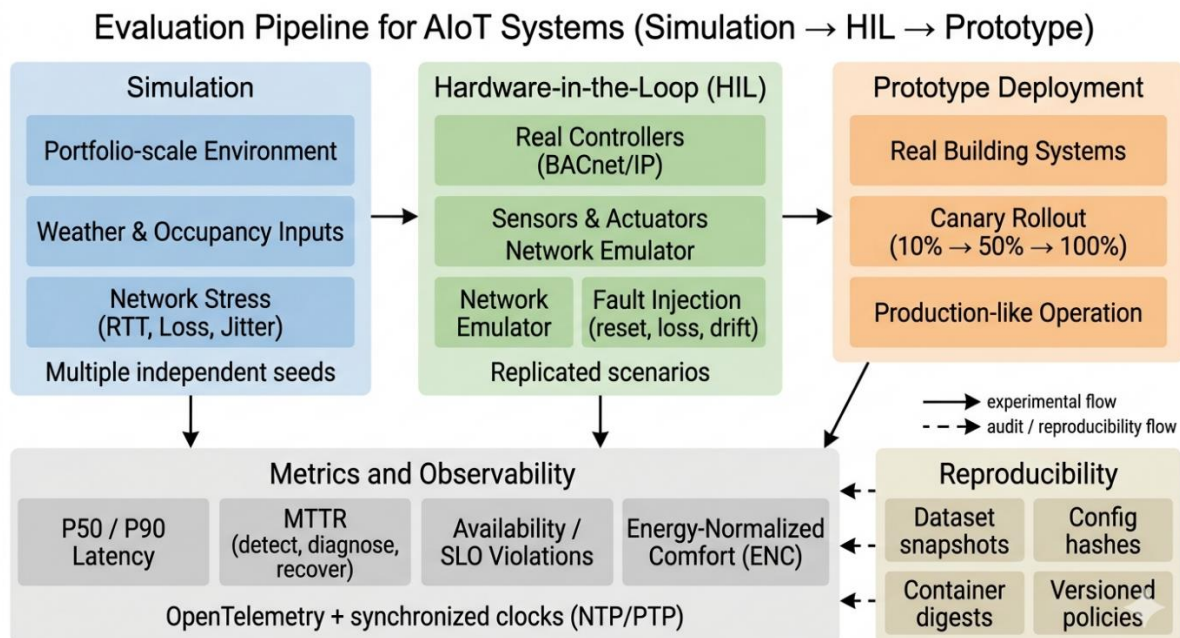


Figure 3. Stage-consistent evaluation pipeline. Simulation, HIL and prototype deployment share metrics, observability and reproducibility artifacts, enabling comparison across increasingly realistic environments.

Table 2. Evaluation stages and measurement focus.

Stage	Environment	Primary stressors	Measured outcomes
Simulation	Portfolio-scale environment	RTT, packet loss, jitter, stochastic occupancy	P50/P90 latency, SLO violations, ENC
HIL	Real BACnet/IP controllers and sensors	Fault injection, device resets, sensor drift	MTTR phases, availability, rollback behavior
Prototype	Production-like building deployment	Canary rollout and realistic schedules	Release cadence, overhead, audit traceability

Results

Across simulation, HIL and prototype stages, the monolithic-modular architecture with governance gates consistently outperforms cloud-centric and microservice baselines. The largest gains appear in tail latency and recovery, where deterministic orchestration and coordinated rollback directly reduce queuing uncertainty and diagnostic delay.

P90 action latency is reduced by approximately 40% relative to the cloud-centric baseline and by a substantial margin relative to the microservice baseline. The improvement remains visible in all three stages, indicating that the effect is not limited to simulation. Median P50 latency also decreases, but the stronger result is the contraction of the tail, which is more relevant for cyber-physical actuation.

Action Latency Across Evaluation Stages

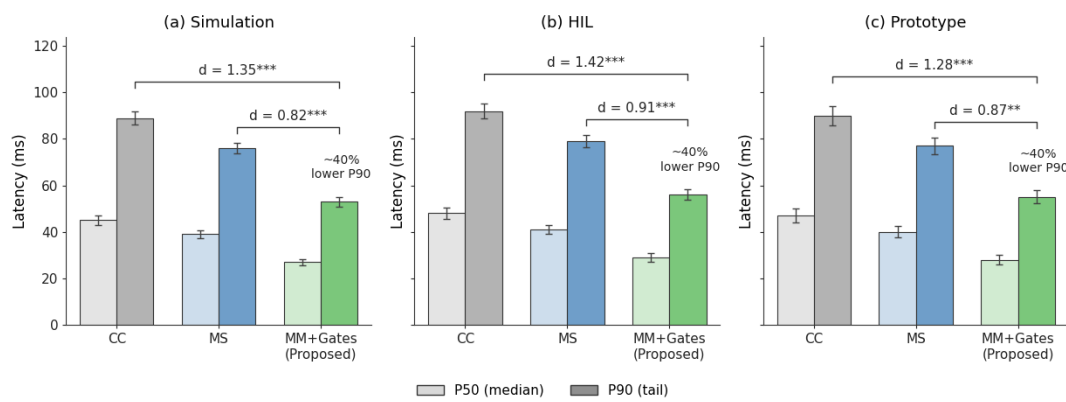


Figure 4. Action latency across simulation, HIL and prototype stages. The proposed MM+Gates architecture shows lower P50 and P90 latency than cloud-centric and microservice baselines.

MTTR decreases by approximately 60% because detection, diagnosis and recovery are coordinated through the orchestrator rather than distributed across independent services. The decomposition shows that the largest contribution comes from shorter diagnosis and deterministic rollback. This is consistent with the design goal of keeping failure handling in the orchestrator rather than distributing it across independent modules.

MTTR Decomposition Across Evaluation Stages

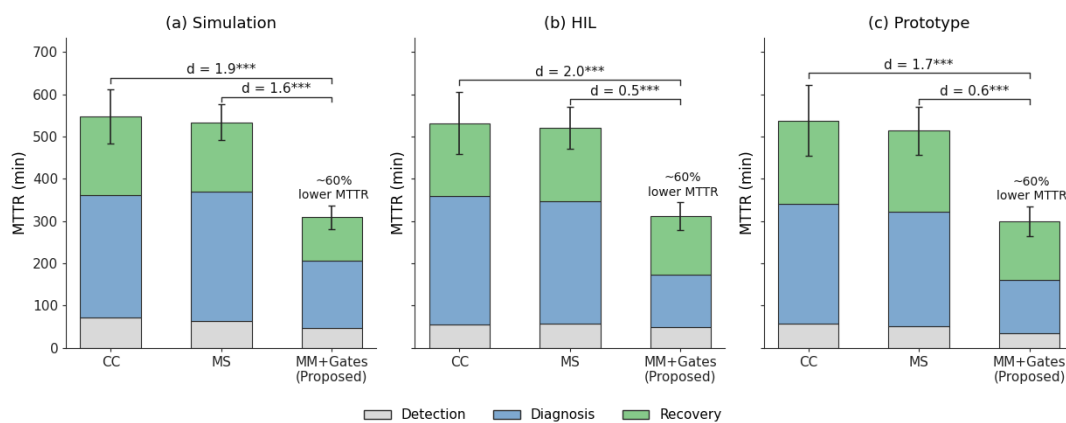


Figure 5. MTTR decomposition across evaluation stages. Coordinated diagnosis and rollback reduce recovery time in simulation, HIL and prototype conditions.

Availability approaches 99.9% under fault injection, with lower availability degradation during repeated disturbance windows. In contrast, cloud-centric and microservice baselines experience deeper dips because recovery is delayed by remote dependencies, distributed diagnosis and less coordinated state restoration.

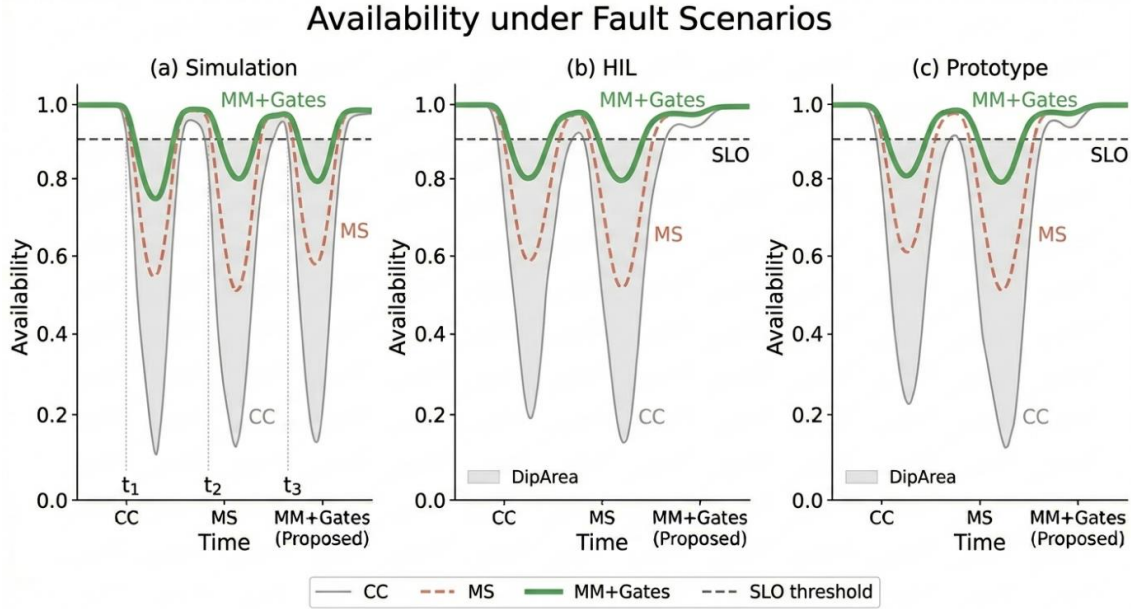


Figure 6. Availability under fault scenarios. The MM+Gates curve remains closer to the SLO threshold and shows lower degradation area across repeated disturbances.

Energy-normalized comfort improves by approximately 15%, indicating that governance and determinism do not prevent energy-aware optimization. The result suggests that a more controlled lifecycle can improve both resilience and sustainability by preventing unstable service versions from reaching production.

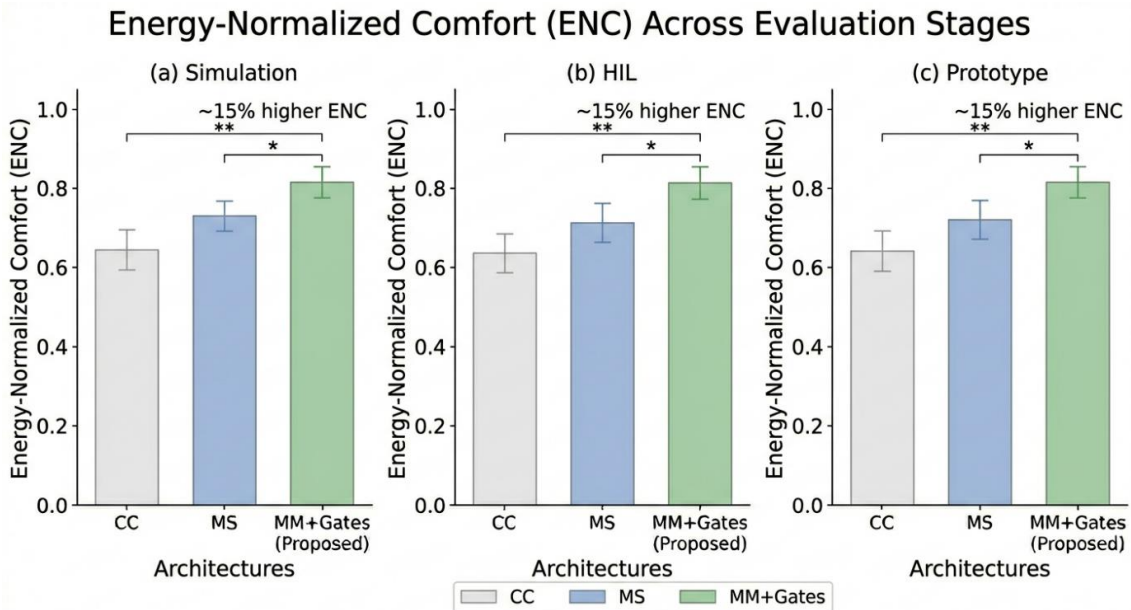


Figure 7. Energy-normalized comfort across evaluation stages. The proposed architecture improves the comfort-energy trade-off in simulation, HIL and prototype deployment.

Governance overhead remains bounded at roughly 3-5%. Although policy evaluation, audit packaging and resilience drills introduce measurable cost, the overhead is offset by faster recovery,

fewer SLO violations and shorter release time. The results therefore support the claim that executable governance can improve operational efficiency rather than merely add compliance burden.

Table 3. Condensed quantitative outcomes.

Metric	Cloud-centric	Microservices	MM+Gates	Improvement
P90 latency (ms)	90-92	76-80	53-56	about 40% lower vs CC
P50 latency (ms)	45-48	39-41	27-30	about 35% lower vs CC
MTTR (min)	520-550	500-530	300-310	about 60% lower vs CC
Availability (%)	98.5-99.0	99.0-99.3	about 99.9	+0.6 to +1.4 pp
ENC score	0.68-0.72	0.72-0.75	0.78-0.82	about 15% higher
Release time (min)	15-18	14-16	10-12	about 30% shorter
Governance overhead	-	-	3-5%	bounded

Multi-objective Trade-off Analysis of AIoT Architectures

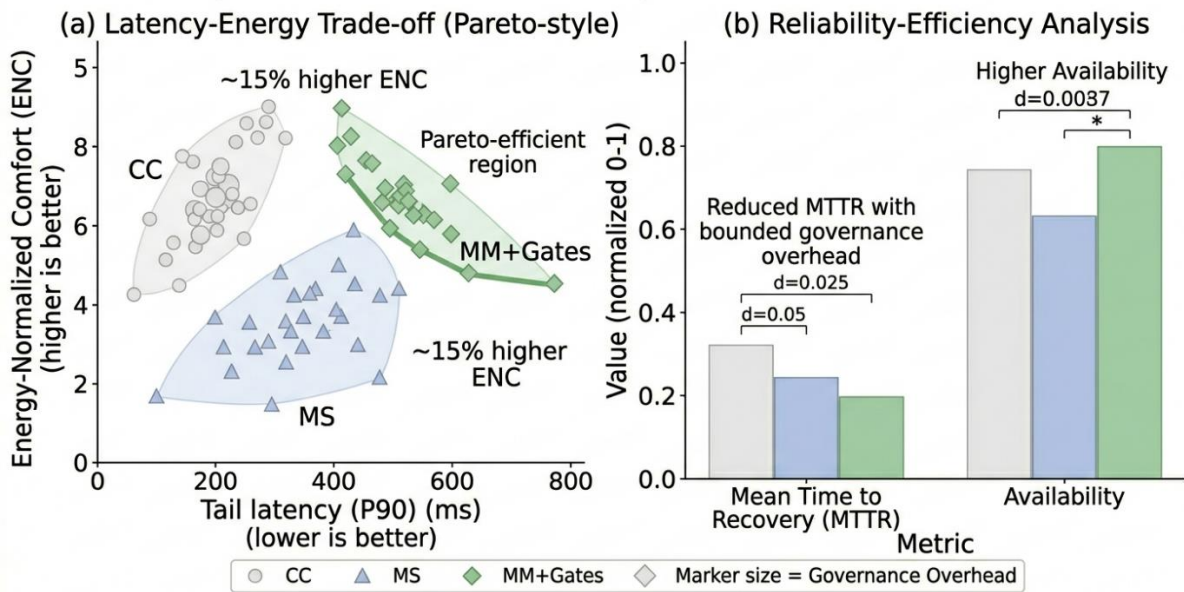


Figure 8. Multi-objective trade-off analysis. The proposed architecture occupies a more favorable region by combining lower tail latency, reduced MTTR, higher availability and improved energy-normalized comfort.

Table 4. Cross-stage meta-analysis and operational overheads.

Outcome	Pooled effect	95% CI	p-value	Interpretation
P90 latency	-35.8 ms	[-38.6, -33.1]	<0.001	Large tail-latency reduction
MTTR	-228.4 min	[-250.7, -206.2]	<0.001	Faster recovery
Availability	+0.92 pp	[+0.74, +1.10]	<0.001	Higher uptime

ENC	+0.11	[+0.08, +0.14]	<0.001	Better comfort-energy score
Governance overhead	3-5%	bounded	-	Operationally acceptable

Discussion

The results show that the main benefit of the framework comes from combining architectural determinism with executable governance. Centralized orchestration bounds fan-out effects and provides a single location for admission, scheduling and recovery decisions. Modular service contracts preserve development agility while preventing unverified modules from directly affecting actuators.

Governance-as-code changes the role of compliance. Instead of collecting evidence after a release, the proposed lifecycle produces evidence during development, policy evaluation, staged rollout and runtime monitoring. This supports auditability and reproducibility because each system state is linked to artifacts, configurations and policy outcomes.

There are limitations. The evaluation includes simulation, HIL and a production-like prototype, but broader external validity requires multi-building and multi-region deployments. Device diversity is also limited compared with real building portfolios. Finally, policy predicates require careful design; weak or incomplete policies would reduce the benefits of the framework. Future work should explore adaptive governance, federated orchestration and open benchmarks that capture both control and lifecycle metrics.

Conclusions

This article presented a shortened journal version of a governance-as-code AIoT framework for smart buildings. The framework integrates deterministic centralized orchestration, modular AI service contracts, normalized device abstraction and executable policy enforcement across CI/CD and runtime operation.

Stage-consistent evaluation shows approximately 40% lower P90 latency, 60% lower MTTR, near-99.9% availability under fault injection, 15% higher energy-normalized comfort and bounded governance overhead. These results indicate that trustworthy AIoT deployment can be treated as an operational systems problem in which governance, reliability and reproducibility are enforced continuously rather than documented separately. The framework provides a practical basis for sustainable, human-centric and auditable AIoT systems in Industry 5.0 environments.

Abbreviations

Abbreviation	Meaning
AIoT	Artificial Intelligence of Things
CI/CD	Continuous integration and continuous delivery
ENC	Energy-normalized comfort
HIL	Hardware-in-the-loop
MTTR	Mean time to repair
P50/P90	Median and tail latency percentiles
SLO	Service-level objective

OPA	Open Policy Agent
-----	-------------------

Reference

1. NIST. AI Risk Management Framework (AI RMF 1.0); NIST: Gaithersburg, MD, USA, 2023.
2. ISO/IEC 42001:2023. Information Technology-Artificial Intelligence-Management System; ISO: Geneva, Switzerland, 2023.
3. IEA. Buildings: Tracking Clean Energy Progress 2024; IEA: Paris, France, 2024.
4. UNEP. 2022 Global Status Report for Buildings and Construction; UNEP: Nairobi, Kenya, 2022.
5. Shi, W.; Cao, J.; Zhang, Q.; Li, Y.; Xu, L. Edge computing: Vision and challenges. IEEE Internet Things J. 2016, 3, 637-646.
6. Dean, J.; Barroso, L.A. The tail at scale. Commun. ACM 2013, 56, 74-80.
7. Richardson, C. Microservices Patterns: With Examples in Java; Manning: Shelter Island, NY, USA, 2018.
8. Beyer, B.; Jones, C.; Petoff, J.; Murphy, N.R. Site Reliability Engineering: How Google Runs Production Systems; O'Reilly Media, 2016.
9. ETSI. Zero-Touch Service Management (ZSM): Reference Architecture Release 2; ETSI GS ZSM 002 V3.1.1, 2020.
10. Open Policy Agent. OPA Gatekeeper Documentation v3.10, 2023.
11. BOPTEST Project. Building Optimization Testing Framework v3.2, 2024.
12. Brick Schema Consortium. Brick Schema 1.3 Release Notes, 2023.
13. RealEstateCore Consortium. RealEstateCore v4.0, 2024.
14. Pineau, J.; Vincent-Lamarre, P.; Sinha, K.; Lariviere, V.; et al. Improving reproducibility in machine learning research. J. Mach. Learn. Res. 2023, 24, 1-20.
15. Attia, A.; Du, Y.; Ahamer, G.; Wang, S.; Lee, K. Closed-loop model predictive control for energy-efficient buildings. Appl. Energy 2023, 332, 120456.
16. Vazquez-Canteli, J.; Nagy, Z. Reinforcement learning for demand response: A review of real-life applications in buildings. Energy Build. 2019, 208, 109636.
17. Kairouz, P.; McMahan, H.B.; Avent, B.; et al. Advances and open problems in federated learning. Found. Trends Mach. Learn. 2021, 14, 1-210.
18. McMahan, B.; Moore, E.; Ramage, D.; Hampson, S.; y Arcas, B.A. Communication-efficient learning of deep networks from decentralized data. AISTATS, 2017.
19. Wen, Q.; Zhou, T.; Zhang, C.; Chen, W.; Ma, Z.; Yan, J.; Sun, L. Transformers in time series: A survey. IJCAI, 2022.
20. IEA EBC Annex 81. Data-Driven Smart Building Operation and Maintenance, 2023.
21. Balasubramanian, V.; et al. Towards a shared and interoperable digital twin for buildings. Springer, 2023.
22. Zhang, Y.; et al. Reproducibility and benchmarking test methods for building energy forecasting models. Appl. Energy 2022, 321, 119307.

Optimizing Educational Processes With Generative AI: Innovative Instructional Design and Automated Assessment Tools

Mukhamadiyev Abdinabi

Research Professor, Department of Computer Engineering, Gachon University, Sujeong-gu, Seongnam-si 13120, Korea

e-mail: mukhamadiyev@gachon.ac.kr

Majidova Yulduz

Assistant, University of Tashkent for Applied Sciences, Department of "Computer Engineering"

e-mail: doniyorovna990@gmail.com

Abstract The rapid emergence of generative artificial intelligence (GenAI) has introduced a structural shift in how educational processes are designed, delivered, and evaluated. This paper examines the role of GenAI in optimizing modern education through three interrelated dimensions: innovative instructional design, automated assessment, and personalized learning. Adopting a structured review and comparative analysis methodology, the study synthesizes recent literature (2021-2026) and documented institutional case studies to identify the mechanisms by which large language models and related technologies enhance teaching efficiency, scalability, and learner engagement. The analysis indicates that GenAI-enabled systems can substantially reduce instructional workload, accelerate content authoring, and provide continuous formative feedback, while raising critical concerns regarding academic integrity, algorithmic bias, data privacy, and over-reliance. A conceptual workflow integrating instructional design and automated assessment within a continuous feedback loop is proposed. The paper concludes that the responsible, human-centered integration of GenAI - supported by educator training, transparent ethical policies, and continuous evaluation - constitutes a strategic competence for contemporary higher education institutions.

Keywords: generative artificial intelligence, instructional design, automated assessment, personalized learning, intelligent tutoring systems, educational technology, large language models, learning analytics.

Introduction

Education systems worldwide are undergoing a profound digital transformation driven by advances in artificial intelligence (AI). Among these advances, generative artificial intelligence - a class of machine-learning systems capable of producing original text, images, code, and multimodal content - has rapidly become one of the most influential technologies in contemporary teaching and learning [1], [2]. Unlike earlier rule-based educational software, generative models learn statistical patterns from large corpora and synthesize context-aware responses, enabling new forms of interaction between learners, educators, and digital systems.

The accelerating adoption of these tools reflects a broader shift toward data-driven, adaptive learning environments. Surveys of higher education indicate that a growing share of institutions are piloting or formally integrating GenAI into curricula, instructional design, and assessment workflows [3], [4]. This trend intensified markedly following the public release of large-scale conversational models in late 2022, which transformed GenAI from a specialized research instrument into a widely accessible educational resource.

Despite this momentum, the integration of GenAI into education remains uneven and contested. While proponents emphasize gains in efficiency, personalization, and accessibility, critics highlight risks to academic integrity, equity, and the development of independent critical thinking [5], [6]. The central problem addressed in this paper is therefore not whether GenAI will influence education, but how its capabilities can be harnessed responsibly to optimize educational processes without compromising pedagogical and ethical standards.

Accordingly, this study pursues three objectives: (i) to characterize the principal applications of GenAI in instructional design, assessment, and personalized learning; (ii) to compare AI-enhanced and traditional educational processes across key effectiveness dimensions; and (iii) to articulate a structured, human-centered framework for the responsible deployment of GenAI in higher education. The remainder of the paper is organized following the IMRAD convention: a review of recent literature, a description of the methodology, a presentation and discussion of results, and concluding recommendations.

Literature Review

Research on artificial intelligence in education (AIED) has expanded from early intelligent tutoring systems toward contemporary generative paradigms. Systematic reviews of the field document a steady diversification of applications, including adaptive learning, profiling and prediction, intelligent tutoring, and automated assessment [3], [7]. These syntheses consistently note that, although technical capability has advanced rapidly, rigorous pedagogical evaluation and ethical reflection have lagged behind.

Generative AI and instructional design. A growing body of work examines how large language models support the authoring of learning materials, lesson plans, and curricula. Studies report that GenAI can accelerate content creation and enable rapid prototyping of differentiated instructional resources [2], [8]. The capacity to generate explanations, worked examples, and assessment items on demand reframes the instructional designer's role from manual production toward curation, validation, and pedagogical alignment.

Intelligent tutoring and personalized learning. Conversational agents and AI tutors have been studied extensively as mechanisms for scalable, individualized support [9], [10]. Recent evidence suggests that dialogue-based systems can provide timely, context-sensitive guidance and sustain learner engagement, particularly when designed to scaffold reasoning rather than supply direct answers. Personalization is frequently identified as the most distinctive contribution of

GenAI, since adaptive systems can tailor pacing, difficulty, and feedback to individual learners at scale [4], [11].

Automated assessment. The literature on automated assessment encompasses automatic grading, formative feedback, and plagiarism or authorship analysis [7], [12]. While automated scoring of structured tasks is comparatively mature, the assessment of open-ended responses using generative models is an active research frontier, raising questions of reliability, fairness, and the appropriate division of labor between human and machine evaluators.

Ethical and practical challenges. A substantial and rapidly growing strand of research addresses the risks associated with GenAI in education. Recurrent themes include threats to academic integrity, the propagation of bias embedded in training data, data-privacy and security concerns, and the danger of learner over-reliance [5], [6], [13]. Meta-level reviews call for stronger ethical frameworks, greater methodological rigor, and closer collaboration between technologists and educators [14]. Institutional guidance, including that issued by international organizations, increasingly emphasizes human oversight, transparency, and equity as preconditions for responsible adoption [1].

Taken together, the literature converges on a dual conclusion: GenAI offers substantial opportunities to optimize educational processes, yet these benefits are contingent on deliberate pedagogical design and robust governance. This paper builds on that consensus by integrating instructional design and automated assessment into a single coherent framework and by comparing AI-enhanced and traditional processes empirically informed by reported trends.

Materials and Methods

This study employs a structured literature review combined with comparative conceptual analysis. The approach is appropriate for a domain that is evolving rapidly and in which consolidated, peer-reviewed evidence coexists with emerging institutional practice.

Data sources and selection. Peer-reviewed articles, systematic reviews, and authoritative institutional reports published between 2021 and 2026 were examined. Sources were identified through academic databases and reference tracing using combinations of search terms such as "generative AI in education," "instructional design," "automated assessment," "intelligent tutoring," and "personalized learning." Priority was given to systematic reviews, high-citation studies, and documents from recognized international bodies to ensure conceptual reliability.

Analytical procedure. The reviewed material was organized thematically into three analytical categories - instructional design, automated assessment, and personalized learning - together with a cross-cutting category of ethical and practical considerations. For the comparative component, five effectiveness dimensions frequently cited in the literature were selected: time efficiency, personalization, scalability, accessibility, and data-driven decision-making. Traditional and AI-enhanced processes were rated on an illustrative 0-100 relative effectiveness index synthesized from reported qualitative and quantitative findings; these values are interpretive rather than the product of a controlled experiment and are intended to visualize directional differences.

Case selection. Three widely documented real-world implementations were selected to illustrate practical deployment across different educational contexts: an AI tutoring system in school education, an AI-augmented language-learning platform, and a university-wide AI integration partnership. These cases were chosen for their public documentation, scale, and relevance to the three analytical categories.

Proposed framework. Finally, the findings were synthesized into a conceptual workflow (Figure 1) that integrates the instructional design pipeline with an automated assessment layer through a continuous feedback loop, with a generative AI engine serving as the shared computational substrate. The methodological limitations of this design - notably the interpretive nature of the comparative index and the reliance on secondary evidence - are acknowledged in the discussion.

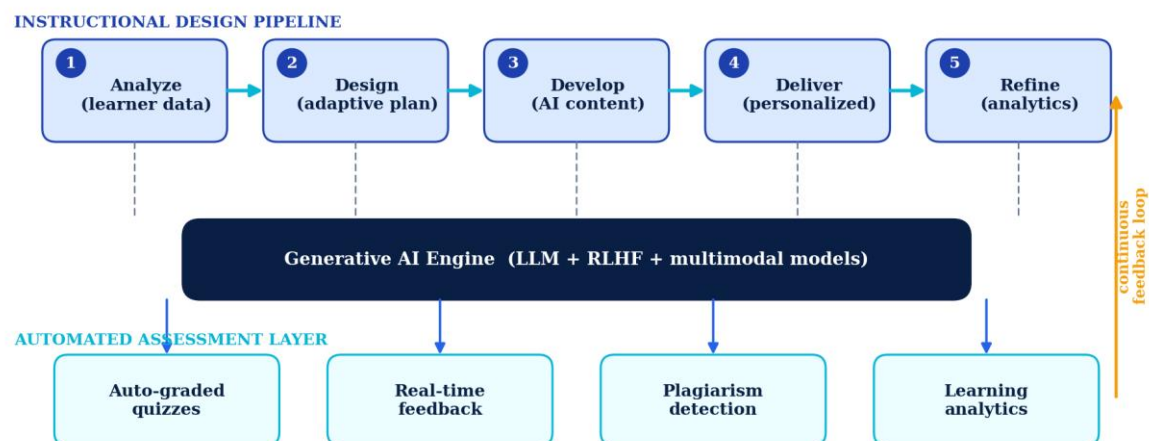


Figure 1. Proposed conceptual workflow integrating the GenAI-based instructional design pipeline (top) with the automated assessment layer (bottom) through a continuous feedback loop. The diagram illustrates how learner data flows through five design stages, is served by a shared generative AI engine, and is continuously refined using assessment analytics.

Results and Discussion

Applications of Generative AI in Education

The thematic analysis confirms that GenAI is being applied across the full educational value chain. Five application areas recur consistently in the reviewed literature: personalized learning, intelligent tutoring, automated content generation, language-learning support, and academic research assistance. Table 1 summarizes these areas, their characteristic functions, and representative technologies.

Application area	Characteristic function	Representative tools
Personalized learning	Adapts pace, difficulty, and feedback to individual learner profiles	Adaptive platforms, LLM tutors
Intelligent tutoring	Provides dialogue-based, context-aware guidance on demand	ChatGPT, Claude, Khanmigo
Automated content generation	Produces lessons, slides, exercises, and worked examples	GenAI authoring assistants
Language-learning support	Enables conversational practice and instant correction	Duolingo Max, speech AI
Research assistance	Summarizes literature and supports citation discovery	LLM research copilots

Table 1. Principal application areas of generative AI in education, with characteristic functions and representative technologies.

Comparative Analysis of Traditional and AI-Enhanced Processes

To assess the optimization potential of GenAI, traditional and AI-enhanced educational processes were compared across five effectiveness dimensions. Table 2 reports the synthesized relative effectiveness indices, and Figure 3 visualizes the same data. Across all dimensions, AI-enhanced processes display a substantial advantage, with the largest gaps observed in personalization and data-driven decision-making.

Effectiveness dimension	Traditional	AI-enhanced	Relative gain
Time efficiency	40	85	+45
Personalization	30	90	+60
Scalability	45	92	+47
Accessibility	50	88	+38
Data-driven decisions	35	95	+60

Table 2. Comparative relative effectiveness index (0-100, illustrative) of traditional versus AI-enhanced educational processes across five dimensions. Values synthesize directional findings from the reviewed literature.

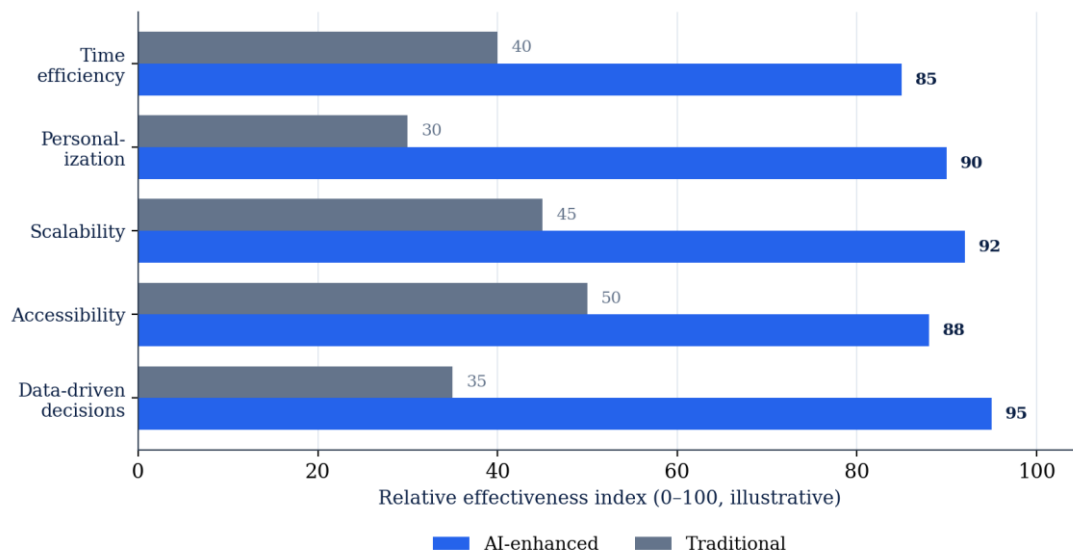


Figure 3. Comparative effectiveness of traditional and AI-enhanced educational processes. AI-enhanced processes consistently outperform traditional approaches, most notably in personalization and data-driven decision-making.

These results align with the literature's emphasis on personalization as the defining advantage of GenAI [4], [11]. The pronounced gains in scalability and time efficiency reflect the capacity of generative systems to automate content authoring and routine assessment, thereby reallocating educator effort toward higher-value pedagogical interaction. The accessibility advantage, while still substantial, is comparatively smaller, underscoring that equitable access depends on infrastructure and policy rather than technology alone.

Adoption Trends

The temporal analysis reveals a steep acceleration in both scholarly attention and institutional adoption of GenAI in education since 2021. Figure 2 illustrates this trajectory, with a marked inflection following the public release of large-scale conversational models in late 2022. Scholarly output and institutional uptake rise in tandem, indicating a closing gap between research interest and practical deployment - a pattern historically uncommon in educational technology, where adoption typically lags research by several years.

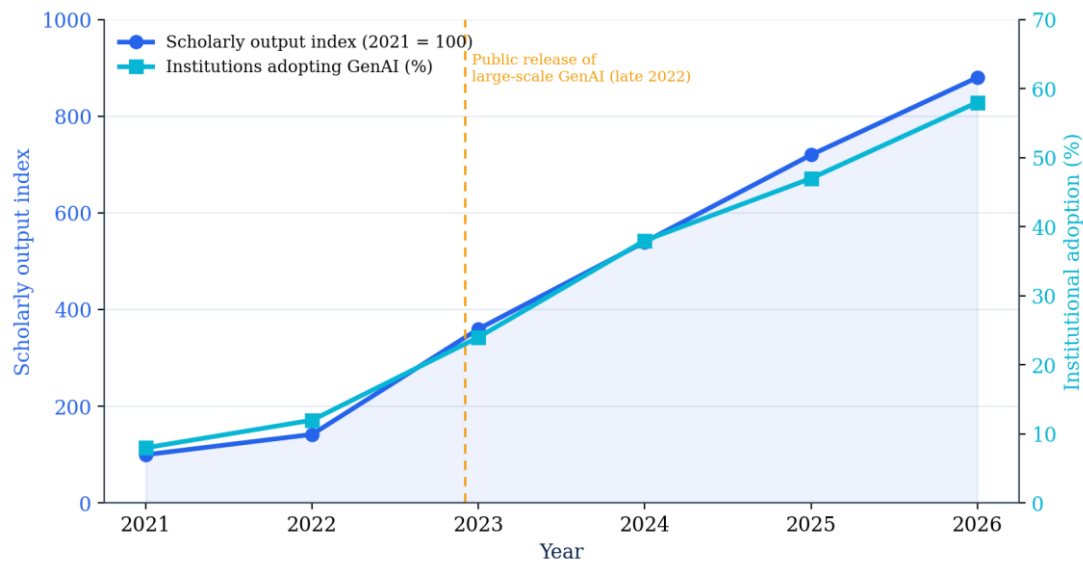


Figure 2. Trend in scholarly output (index, 2021 = 100) and institutional adoption of generative AI in education (2021-2026). The dashed line marks the public release of large-scale GenAI models in late 2022. Values are illustrative and indicate directional momentum.

Real-World Case Studies

Three documented implementations illustrate how the application areas in Table 1 translate into practice. Table 3 summarizes the cases, the GenAI capabilities they leverage, and their educational context.

Case / platform	GenAI capability	Context and reported outcome
Khanmigo (Khan Academy)	Dialogue-based tutoring	K-12 setting; Socratic guidance that scaffolds reasoning rather than supplying answers
Duolingo Max	Conversational language practice	Global users; AI-driven explanation and dialogue features integrated into language learning
University-OpenAI partnership	Campus-wide AI assistants	Higher education; institution-wide deployment for teaching, learning, and research support

Table 3. Representative real-world case studies of generative AI in education, mapped to GenAI capability and educational context.

Collectively, the cases demonstrate that effective deployments share a common design principle: GenAI augments rather than replaces human pedagogy. The tutoring case, in particular, shows that constraining the model to a scaffolding role - guiding learners through questioning - mitigates the risk of dependency while preserving the benefits of on-demand support.

Ethical and Practical Challenges

The optimization potential of GenAI is counterbalanced by significant challenges. Academic integrity is the most frequently cited concern, as generative systems can produce work that is difficult to distinguish from authentic learner output [5], [13]. Algorithmic bias, inherited from training data, threatens fairness across linguistic and cultural groups [6]. Data privacy is a further imperative, given that adaptive systems process sensitive learner information. Finally, over-reliance on AI may erode critical thinking if learners delegate cognitive effort rather than engaging with it. These risks are not arguments against adoption but rather requirements for governance: transparent policies, human oversight, and continuous quality assurance are necessary conditions for responsible use [14].

Future Trends

Looking forward, the literature and observed trajectories point to five converging developments: AI-enabled smart classrooms that integrate the Internet of Things with adaptive systems; AI-powered universities offering institution-wide intelligent services; ubiquitous virtual tutors providing multimodal, subject-agnostic support; deeper human-AI collaboration in which educators direct and validate AI capabilities; and interoperable educational ecosystems unified by intelligent agents. The common thread is augmentation rather than substitution - a future in which AI extends human pedagogical capacity within clearly defined ethical boundaries.

Conclusion

This paper examined the role of generative artificial intelligence in optimizing educational processes through innovative instructional design, automated assessment, and personalized learning. The analysis yields three principal conclusions. First, GenAI is a transformative technology that materially improves time efficiency, personalization, scalability, accessibility, and data-driven decision-making relative to traditional approaches. Second, the most effective implementations are human-centered: GenAI augments educators rather than replacing them, and its value depends on sound pedagogical design. Third, responsible integration - underpinned by educator training, transparent ethical policies, and continuous evaluation - is now a strategic competence for higher education institutions.

On the basis of these findings, the study recommends that institutions adopt explicit AI governance frameworks, invest in educator AI literacy, define clear policies on academic integrity and data privacy, preserve a central role for human judgment, and establish continuous monitoring of educational outcomes. The proposed workflow (Figure 1) offers a practical structure for embedding these principles within everyday instructional and assessment practice.

Future research should move beyond synthesis toward controlled, longitudinal studies that quantify the learning impact of GenAI across diverse contexts, validate automated assessment of open-ended work, and develop robust, equitable governance mechanisms. As generative technologies mature, the decisive question will not be whether AI transforms education, but whether educators and institutions can shape that transformation in a manner that is pedagogically sound, ethical, and inclusive.

Reference

1. UNESCO, Guidance for Generative AI in Education and Research. Paris: UNESCO, 2023.
2. Z. Bahroun, C. Anane, V. Ahmed, and A. Zacca, "Transforming education through generative artificial intelligence: A comprehensive review," *Sustainability*, vol. 15, no. 17, p. 12983, 2023.
3. T. K. F. Chiu, Q. Xia, X. Zhou, C. S. Chai, and M. Cheng, "Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education," *Computers and Education: Artificial Intelligence*, vol. 4, p. 100118, 2023.
4. H. Crompton and D. Burke, "Artificial intelligence in higher education: the state of the field," *International Journal of Educational Technology in Higher Education*, vol. 20, no. 1, p. 22, 2023.
5. E. Kasneci et al., "ChatGPT for good? On opportunities and challenges of large language models for education," *Learning and Individual Differences*, vol. 103, p. 102274, 2023.
6. L. Yan et al., "Practical and ethical challenges of large language models in education: A systematic scoping review," *British Journal of Educational Technology*, vol. 55, no. 1, pp. 90-112, 2024.
7. O. Zawacki-Richter, V. I. Marin, M. Bond, and F. Gouverneur, "Systematic review of research on artificial intelligence applications in higher education," *International Journal of Educational Technology in Higher Education*, vol. 16, no. 1, p. 39, 2019.
8. D. Baidoo-Anu and L. O. Ansah, "Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning," *Journal of AI*, vol. 7, no. 1, pp. 52-62, 2023.
9. M. A. Kuhail, N. Alturki, S. Alramlawi, and K. Alhejori, "Interacting with educational chatbots: A systematic review," *Education and Information Technologies*, vol. 28, no. 1, pp. 973-1018, 2023.
10. A. Tlili et al., "What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education," *Smart Learning Environments*, vol. 10, no. 1, p. 15, 2023.
11. C. K. Lo, "What is the impact of ChatGPT on education? A rapid review of the literature," *Education Sciences*, vol. 13, no. 4, p. 410, 2023.
12. Y. K. Dwivedi et al., "Opinion paper: 'So what if ChatGPT wrote it?' Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI," *International Journal of Information Management*, vol. 71, p. 102642, 2023.
13. D. Mhlana, "Open AI in education, the responsible and ethical use of ChatGPT towards lifelong learning," in *FinTech and Artificial Intelligence for Sustainable Development*. Cham: Springer, 2023, pp. 387-409.
14. M. Bond et al., "A meta systematic review of artificial intelligence in higher education: a call for increased ethics, collaboration, and rigour," *International Journal of Educational Technology in Higher Education*, vol. 21, no. 1, p. 4, 2024.
15. OECD, *OECD Digital Education Outlook 2023: Towards an Effective Digital Education Ecosystem*. Paris: OECD Publishing, 2023.
16. EDUCAUSE, *2024 EDUCAUSE Horizon Report: Teaching and Learning Edition*. Boulder, CO: EDUCAUSE, 2024.

17. G. Cooper, "Examining science education in ChatGPT: An exploratory study of generative artificial intelligence," *Journal of Science Education and Technology*, vol. 32, no. 3, pp. 444-452, 2023.
18. S. Grassini, "Shaping the future of education: Exploring the potential and consequences of AI and ChatGPT in educational settings," *Education Sciences*, vol. 13, no. 7, p. 692, 2023.
19. W. Holmes and I. Tuomi, "State of the art and practice in AI in education," *European Journal of Education*, vol. 57, no. 4, pp. 542-570, 2022.
20. J. Su and W. Yang, "Unlocking the power of ChatGPT: A framework for applying generative AI in education," *ECNU Review of Education*, vol. 6, no. 3, pp. 355-366, 2023.

A Speech-Driven Framework for Automated Uzbek-Language Customer Service

Mukhamadiyev Abdinabi

Research Professor, Department of Computer Engineering, Gachon University, Sujeong-gu, Seongnam-si 13120, Korea

e-mail: mukhamadiyev@gachon.ac.kr  **orcid** 0000-0002-1438-0628

Khujakulov Toshtemir

University of Tashkent for Applied Sciences, Department of "Computer Engineering" t.f.f.D (PhD), Associate Professor

e-mail: mehr_toj@mail.ru  **orcid** 0000-0002-4906-9258

Majidova Yulduz

Assistant, University of Tashkent for Applied Sciences, Department of "Computer Engineering"

e-mail: doniyorovna990@gmail.com

Abstract Contact centers in banking and public services increasingly struggle to combine continuous availability with acceptable waiting times, and this problem is especially acute in low-resource-language markets such as Uzbekistan, where mature commercial speech technology is unavailable. To address this gap, we design, implement, and evaluate UzAssistant, a fully speech-driven virtual operator that conducts an entire service dialogue in Uzbek without human intervention. The framework chains four neural components. First, a Gaussian-Mixture-Model classifier operating on Mel-Frequency Cepstral Coefficient (MFCC) features verifies the caller's identity against enrolled voice profiles, with a rejection threshold for unregistered speakers. Second, an end-to-end recurrent architecture based on Deep Speech 2 transcribes the spoken request into standardized Uzbek text. Third, an attention-augmented sequence-to-sequence summarizer condenses the transcription, after which a fine-tuned multilingual sentence transformer embeds the condensed query and selects the semantically closest entry from a structured question-answer knowledge base through cosine similarity. Finally, the retrieved answer is vocalized by a WaveNet neural vocoder coupled with a Tacotron 2 decoder and explicit phoneme alignment, producing natural Uzbek speech in real time. Trained and evaluated on a purpose-built 108-hour banking speech corpus of 863 speakers, the framework reached 96.4% speaker-verification accuracy with 56 live participants, 85.27% average sentence-embedding accuracy over four Uzbek datasets, and 82.5% end-to-end response accuracy on paraphrased customer queries; the strongest recognition configuration-Deep Speech 2 with language-model decoding and threefold data augmentation-yielded a 5.22% character error rate and a 13.8% word error rate. These findings indicate that operator workload in Uzbek-language call centers can be reduced by an estimated 20-25% while extending service to a 24/7 schedule.

Keywords: speech technologies; call center; speech corpus; Uzbek language; speech-to-text; text-to-speech; speech recognition; speech synthesis; IVR; public services

Introduction

Research Context and Motivation

The rapid rise of artificial intelligence (AI) has been accompanied by the widespread adoption of voice assistants (VAs) such as Amazon Echo, Google Assistant, Microsoft Cortana, and Apple Siri. These AI-driven assistants are reshaping how people interact with technology: industry analyses report that roughly 27% of internet users worldwide employ voice search [2], and forecasts predicted a doubling of in-home voice-assistant use between 2018 and 2023 [3]. Many experts anticipate that voice assistants will complement conventional computing devices, particularly for practical tasks such as shopping [4]. Although concerns about privacy and security persist, advances in natural language processing (NLP) and machine learning now allow such assistants to sustain complex conversations and manage multiple tasks simultaneously [5], suggesting profound changes in future human-machine interaction.

Automatic speech recognition (ASR) and speech synthesis are already central to call-center applications, including automatic call distribution (ACD), interactive voice response (IVR) systems, and workforce-management tools. These technologies enable intelligent voice menus for incoming calls, automated outbound calling with a voice interface, transcription of telephone conversations, real-time recommendations for operators, recognition of customer emotions, and systematic evaluation of service quality. Nevertheless, most mature ASR and TTS solutions target high-resource languages such as English or Russian [13-16], leaving speakers of low-resource languages, including Uzbek, underserved. Because call-center communication takes the form of live conversation, speech recognition must operate in real time or near-real time, which raises additional engineering challenges [17-20].

Deploying speech technologies in call centers offers several concrete operational benefits [21-25]: substantial reductions in average handling time and associated labor cost; a 1.5-2 $\times$ reduction in call duration by eliminating manual data entry; automatic handling of simple, repetitive questions so that operators can concentrate on complex calls; uninterrupted 24/7 service, including holidays; voice-based caller verification, which is particularly valuable in banking for protecting cards and confidential documents; the capacity to absorb large volumes of short calls; replacement of error-prone tone-mode IVR menus; and the use of recognized speech as an additional analytical resource for improving first-call resolution (FCR), thereby lowering callback rates and operating costs while increasing customer satisfaction.

Research Aims and Contributions

Our research indicates that combining complementary neural architectures can markedly improve the precision of automated customer-service systems for the Uzbek language and its dialects. By employing RNN encoder-decoder, DNN-CTC, end-to-end Transformer, and end-to-end Conformer models, both statistical and neural language models can be constructed within ASR frameworks that accurately identify customer voice requests and deliver responses tailored to the Uzbek context [6-12]. This study aims to close the gap between academic speech research and practical call-center deployment for a low-resource language. The key contributions are as follows:

1. Speaker recognition in varied environments: a GMM-based speaker-recognition module was evaluated in both noisy and studio-quality conditions, achieving 96.4% accuracy in real-time tests with 56 participants.
2. Feature extraction with Deep Speech 2: the Deep Speech 2 architecture was employed for extracting MFCC features from user utterances.
3. Uzbek end-to-end ASR: an RNN-based end-to-end ASR model was trained on a large annotated Uzbek speech corpus to convert consumer speech into text.
4. Sentence summarization: a Seq2Seq model with attention condenses customer utterances, and a BERT sentence transformer embeds sentences and search queries with an average accuracy of 85.27%.
5. Structured database design: three coordinated databases-a Personal Information Database (PID), a Generic Information Database (GID), and a Credential Information Database (CID)-manage user data, FAQs, and sensitive credentials.
6. A natural-voice Uzbek speech synthesizer for call centers: a WaveNet-based TTS engine, calibrated to the phonetic characteristics of Uzbek, initiates dialogue with callers and gathers the required information, reducing the preliminary conversational stages traditionally handled by human operators.
7. Analysis of practical deployment challenges: the study examines the linguistic variation, background noise, and integration issues that arise when speech technology is embedded in public-service call centers.
8. Efficiency projections: preliminary analysis suggests that judicious deployment of automated speech recognition and synthesis could reduce manual operator involvement by 20-25% while enabling continuous 24/7 customer service.

Related Work

AI-driven solutions have proliferated in business operations in recent years [26], introducing cognitive capabilities such as automation, image recognition, problem solving, and informed decision making [27]. These capabilities are delivered through chatbots, intelligent virtual interfaces, robotic systems, and related digital tools [28], which raise productivity across education, healthcare, management, and industrial production [27,29,30]. AI's capacity to process and represent complex data quickly improves organizational efficiency and simplifies decision making [31,32], although growing dependence on intelligent machines has also prompted debate about ethical and social consequences [33].

Neural language modeling for speech recognition has a long lineage. Schwenk and Gauvain [34] pioneered artificial neural network (ANN) language models, contrasting an ANN-driven N-gram model with Kneser-Ney smoothing on a corpus exceeding 550 million words and reducing word misrecognition by 0.5% through neural rescoring. Mikolov et al. [35] introduced RNN-based language modeling with frequency-based word classing; rescoring the top-100 hypotheses of a 5-gram baseline with the RNN model cut the word error rate (WER) by 18% while reducing model complexity. Huang et al. [36] applied an RNN language model in the first decoding pass of Bing voice search, adding a key-value hash cache for efficiency and lowering WER from 25.3% to 23.2%; combining lattice generation with a 4-gram model and RNN-based rescoring further reduced

WER to 22.7%. Sundermeyer et al. [37] systematically compared feedforward and recurrent neural language models trained on a 27-million-word corpus, observing consistent WER reductions of 1.4-1.5% and a 0.4% advantage of RNNs over feedforward ANNs. Morioka et al. [38] proposed a variable-length-context language model that reduced both perplexity and WER in large-vocabulary recognition tests.

In the dialogue-systems domain, Hardy et al. [39] proposed a dialogue-management framework integrating ASR, TTS, a dialogue manager, and a large information database for telephone-based automated customer service. Zweig et al. [40] introduced a quality-monitoring system for call centers that combines speaker recognition, maximum-entropy classification, pattern recognition, and ASR. McLean and Osei-Frimpong [41] surveyed 302 participants to identify the determinants of customer satisfaction in live-chat service encounters, and Warnapura et al. [42] designed an AI-based architecture that delivers information to customers via text, voice, and e-mail, combining sentiment analysis, NLP, and ASR.

For the Uzbek language specifically, Mansurov and Mansurov [43] introduced UzBERT, a BERT model pretrained on a news corpus of more than 142 million words with masked language modeling and next-sentence prediction objectives; its architecture mirrors the original BERT-base design (12 layers, 768 hidden units, 12 attention heads, 110 million parameters, 30,000-token vocabulary). Ren et al. [44] proposed multiscale parallel convolution (MSPC) feature extraction with a bidirectional LSTM and a fine-tuned BERT-initialized RNN language model to improve hybrid CTC/attention ASR for agglutinative languages. Mamatov et al. [45] developed a Transformer-based Uzbek speech-recognition system trained on 432 hours of audiobook recordings and 72 hours of proverb recordings from 174 speakers. For sentence-similarity measurement, Doc2Vec [46] and Word2Vec [47] remain widely used neural approaches for capturing the semantic content of words, sentences, and paragraphs.

In summary, although NLP and speech recognition have each received substantial academic attention, their empirical integration into operational call-center environments-especially for low-resource Central Asian languages-remains under-examined. The present study addresses this gap by designing, implementing, and evaluating a complete Uzbek-language voice-assistant pipeline for call-center automation.

Methodology

System Workflow

The proposed system, UzAssistant, is organized into four principal components: (i) speech recognition, (ii) sentence summarization, (iii) sentence-similarity measurement, and (iv) text-to-speech conversion. When a customer call arrives, the audio is first passed to the speaker-recognition and ASR modules, which identify the caller and transcribe the utterance into Uzbek text. The transcription is then summarized by a Seq2Seq model with attention to distill the essential content of the request. The summarized query is embedded and compared against a curated question-answer dataset using a sentence-similarity module, and the best-matching answer is generated in text form. Finally, the WaveNet-based TTS engine converts the textual answer into natural-sounding Uzbek speech, which is played back to the caller. Figure 1 presents a schematic representation of the complete pipeline.

TTS technology simulates human-like speech by transforming written text into audio using machine-learning methods, and it is particularly valuable for voice-operated robots and IVR systems because it eliminates manual audio recording and editing. Two families of approaches dominate: concatenative TTS, which stitches together pre-recorded audio fragments and delivers high quality at the cost of large training datasets, and parametric TTS, which uses a probabilistic model to predict the acoustic characteristics of the signal from input text and can approach human-like speech with considerably less data [48]. The present work adopts a neural parametric approach based on WaveNet, as detailed in Section 3.5.

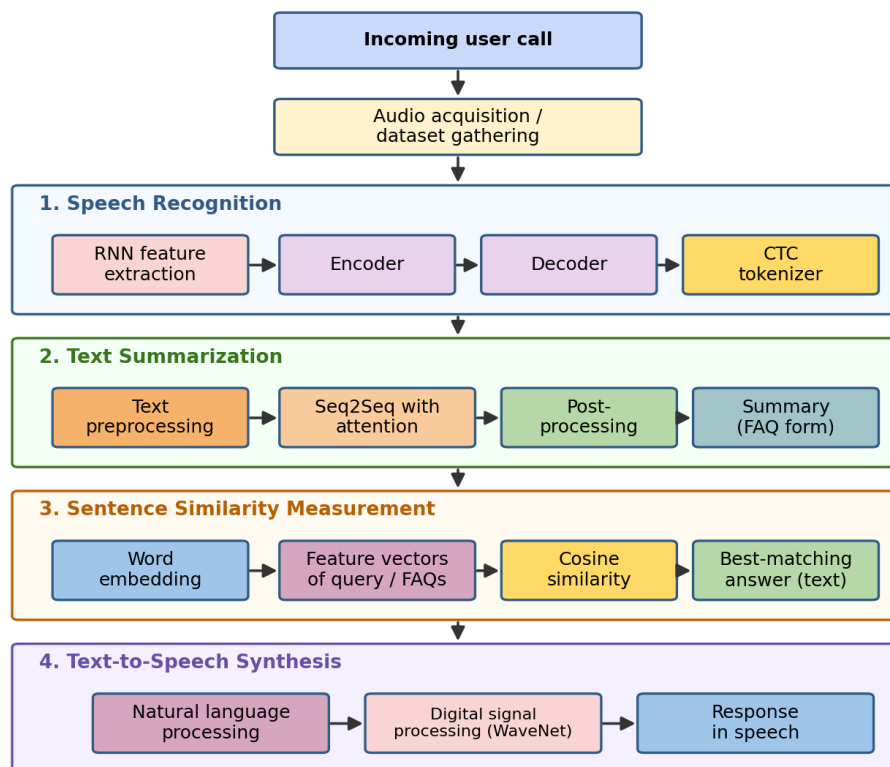


Figure 1. Schematic representation of the UzAssistant components (speech recognition, text summarization, sentence-similarity measurement, and text-to-speech synthesis).

Speaker Recognition

Speaker recognition employs the Mel-Frequency Cepstral Coefficients (MFCC) method [49] to extract features from the audio signal, focusing primarily on low-frequency components, together with a feature-matching technique driven by an expectation-maximization algorithm over Gaussian Mixture Models (GMMs) [50]. For speaker adaptation, the system collects a 20-second voice sample from each user at registration and constructs an individual GMM profile from the extracted features, enabling precise similarity assessment at call time. During identification, the test utterance is scored against all registered speaker models, and the maximum score is compared with a decision threshold: if the score meets or exceeds the threshold, the caller is accepted as the corresponding registered speaker; otherwise, the caller is classified as unknown. This thresholding mechanism

substantially reduces misidentification of unregistered speakers. The registration and identification procedure is depicted in Figure 2.

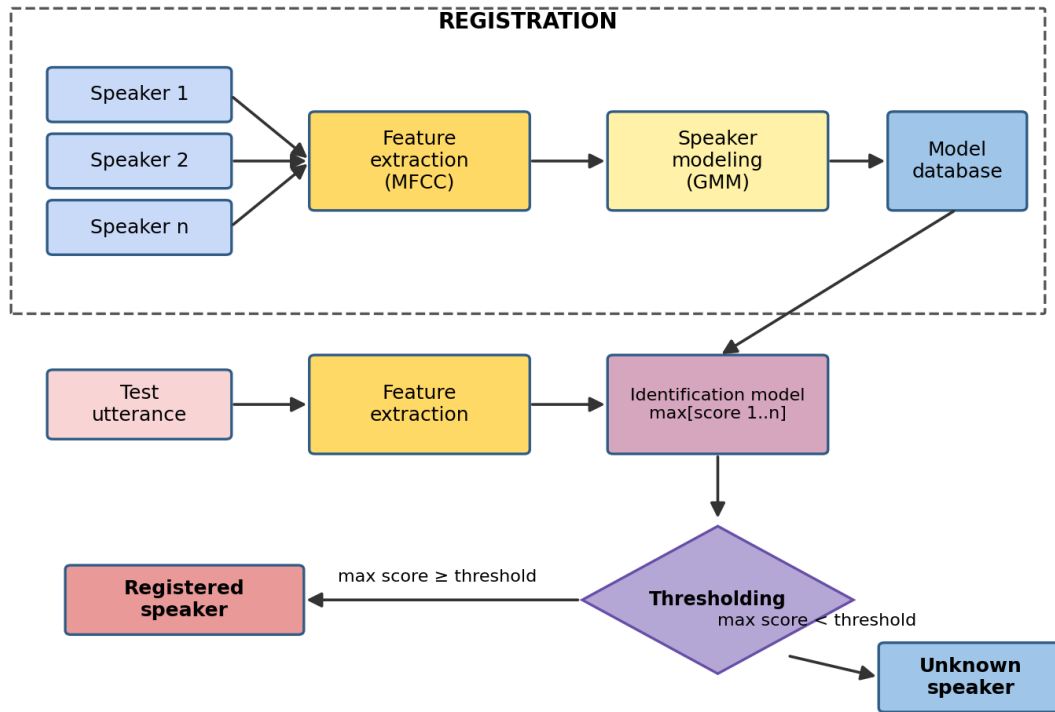


Figure 2. Procedure of the speaker-recognition system: registration (feature extraction and GMM modeling) and identification (scoring against stored models with threshold-based decision).

Automatic Speech Recognition

For accurate and efficient transcription of consumer speech, the ASR model was trained on a publicly available annotated Uzbek voice corpus using the Deep Speech 2 architecture [51]. Deep Speech 2 exploits high-performance-computing techniques and batch normalization to train approximately seven times faster than its predecessor and employs the SortaGrad curriculum-learning strategy [50]. Decoding selects the label sequence that maximizes the connectionist temporal classification (CTC) probability:

$$l^* = \operatorname{argmax}_{l \in \operatorname{Align}(x,y)} \prod_{t=1..T} p_{ctc}(l_t | x; \theta) \quad (1)$$

where the function $\operatorname{Align}(x, y)$ maps every admissible pairing of characters in the transcription y to frames of the input x , thereby synchronizing the text transcription with the input audio frames.

Sentence Summarization

An Uzbek sentence-summarization component was developed to shorten utterances and distill the key information from customer replies. The component is built on the Seq2Seq summarization framework [52] enhanced with an attention mechanism, which reduced the training loss to 0.001 in our experiments. The architecture of the summarization system is shown in Figure 3.

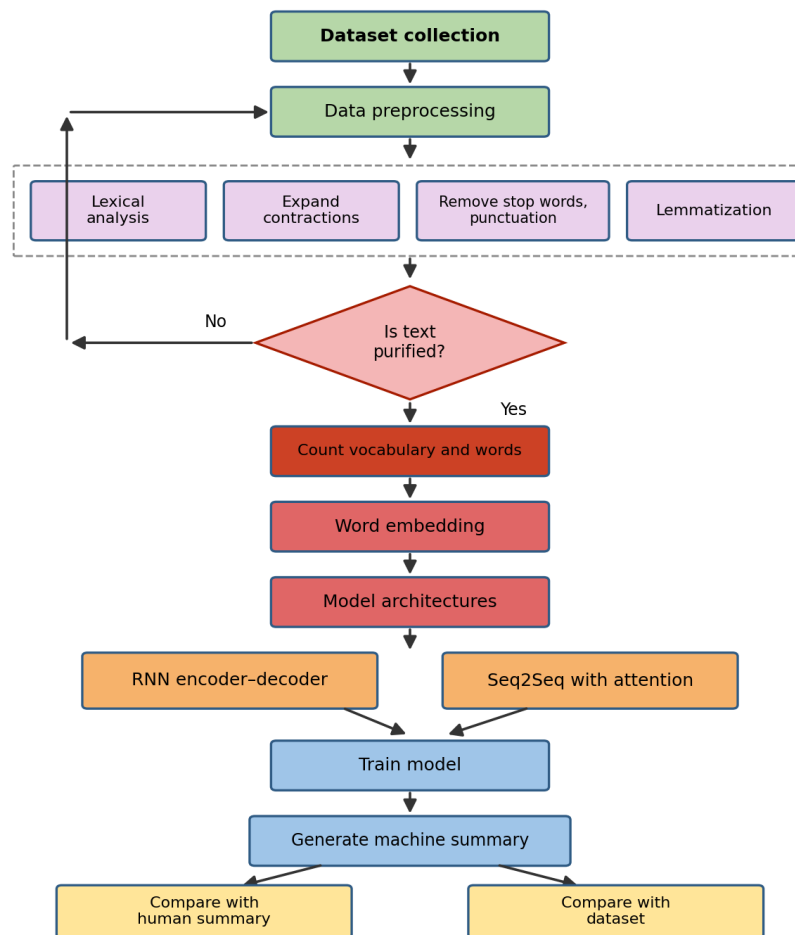


Figure 3. Structure of the sentence-summarization system: preprocessing (lexical analysis, contraction expansion, stop-word removal, lemmatization), word embedding, RNN encoder-decoder/Seq2Seq-with-attention training, and comparison of machine summaries against human summaries.

Data collection and preprocessing. The Uzbek Bank Speech Corpus (UBSC v1.0) is a comprehensive dataset assembled for speech-to-text conversion. It contains 108 hours of recorded Uzbek speech in WAV format from 863 speakers of diverse ages, genders, dialects, education levels, and accents. The same corpus supports the text-summarization task with 35 k short articles and 24 k short reference summaries, and 322 questions harvested from the City Bank website and social-network pages serve as the sentence-similarity evaluation set. Preprocessing proceeds in four steps: sentences are tokenized; contractions are expanded and stop words and punctuation removed; lemmatization reduces words to their base forms; and words are tagged by part of speech. These steps substantially improve the quality of downstream analysis.

Model architecture. Training combines an RNN encoder-decoder with the Seq2Seq approach and an attention mechanism to condense articles effectively. The design comprises three segments-encoder, attention, and decoder. The encoder transforms an input sequence into a fixed context vector that captures the semantic abstraction of the entire article; this vector initializes the decoder's hidden state despite timestamp disparities between encoder and decoder. Denoting the target sentence set by a and the source set by b , the model seeks the word-vector sequence of maximum likelihood:

$$\arg \max_b p(a | b) \quad (2)$$

A sequence-to-sequence framework with Bahdanau attention aligns the fixed-length output. For embedding, a pre-trained Uzbek word-vector model ("uz w2c model") converts words into numeric vectors, and the final vectorized representation of each word is used for model training.

Text-to-Speech Synthesis

Because the system targets the State Services Center sector in Uzbekistan, where the overwhelming majority of users communicate in Uzbek, the TTS module must produce natural Uzbek speech in real time. The synthesis engine is built on DeepMind's WaveNet [53], a deep generative model of raw audio waveforms that, when trained on real speech data, produces highly realistic human-like voices and outperformed the best contemporaneous text-to-speech systems for American English and Mandarin. The WaveNet architecture used in this work is illustrated in Figure 4.

Producing a high-quality synthetic TTS database requires the speech output of the source TTS model to align precisely with phonetic pronunciations. To this end, a Tacotron 2 decoder equipped with a phoneme-alignment method [54] synchronizes phoneme sequences with their corresponding acoustic features. An external duration-prediction model determines the length of each phoneme from linguistic attributes; the Tacotron 2 decoder then produces the acoustic features, which a WaveNet-based neural excitation vocoder converts into speech signals. Within the vocoder, a WaveNet-based mixture-density network [55] operates according to the principles of human speech production [56], yielding stable and accurate speech generation [57,58].

The time-frequency trajectory excitation vocoder extracts features every 5 ms [59]: 40-dimensional line spectral frequencies, the fundamental frequency, energy, a voicing indicator, a slowly evolving 32-dimensional waveform, and a rapidly evolving 4-dimensional waveform-together a 79-dimensional feature vector. The acoustic model of the source TTS comprises a context analyzer, a context encoder, and a Tacotron decoder. The context analyzer extracts 354-dimensional phoneme-level linguistic feature vectors (330 categorical and 24 numerical contexts). A duration predictor with three fully connected layers (1,024, 512, and 256 units) and an LSTM layer with 128 memory blocks estimates phoneme durations, after which phoneme-level features are upsampled to frame level. The context encoder refines the frame-level features through three convolution layers (10×1 kernels, 512 channels), a bidirectional LSTM with 512 memory blocks, and fully connected layers of 512 units. The Tacotron decoder-PreNet, PostNet, and a primary unidirectional LSTM-then generates the acoustic features: the PreNet (two fully connected layers of 256 units) processes previously generated features, whose outputs, together with a context-embedding module, pass through two unidirectional LSTM layers with 1,024 memory blocks and two 79-unit projection layers; the PostNet (five convolution layers, 5×1 kernels, 512 channels) adds residual refinements that sharpen generation precision.

In the WaveNet configuration, dilation factors follow the sequence $[2^0, 2^1, \dots, 2^9]$ repeated three times, producing 30 residual-block layers and a receptive field of 3,067 samples; each residual block uses 128-channel convolution layers. The network outputs two dimensions-the mean and standard deviation of a Gaussian distribution-and weight normalization constrains all weight vectors to a standard length [60]. As post-processing, a spectral-domain sharpening filter with coefficient

0.95 enhances the spectral definition of the synthesized speech, and the WaveNet scale parameter in voiced sections is reduced by a factor of 0.87 to produce clearer audio.

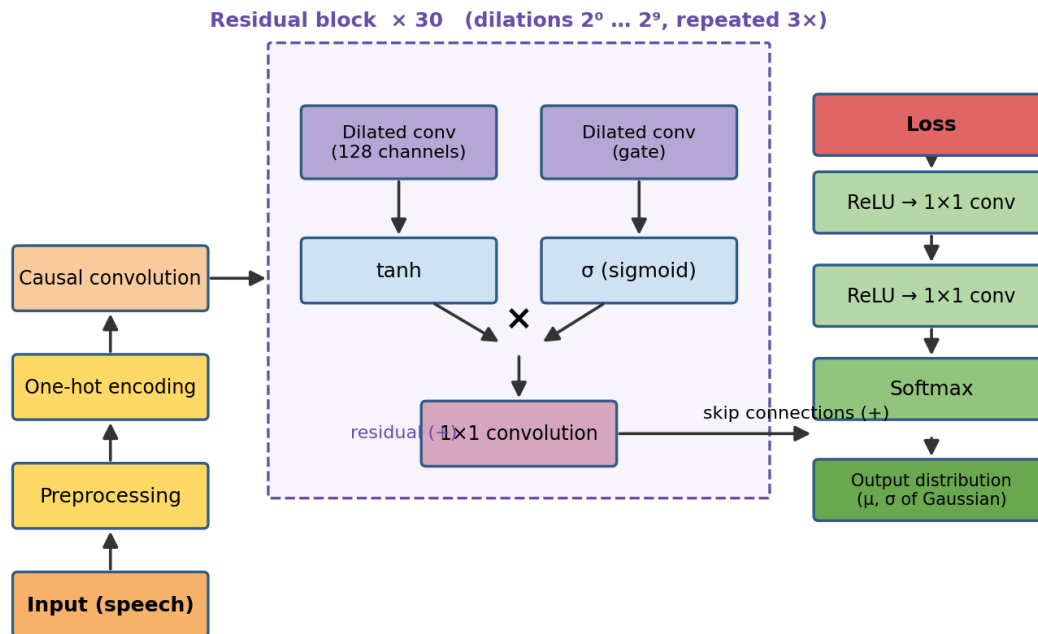


Figure 4. WaveNet architecture with dilated causal convolutions, gated activation units, residual and skip connections, and the softmax output stage.

Database Design

The system relies on three coordinated databases. The Personal Information Database (PID) stores user details-full name, National Identity (NID) number, mobile number, date of birth, the text form of the recorded response, the speaker-specific GMM model, and the call timestamp-and is used to verify callers by cross-referencing submitted personal information (Table 1). The Generic Information Database (GID) houses frequently asked questions for the service domain and is the primary resource for new users (Table 2). The Credential Information Database (CID) safeguards sensitive information and defines the domain-specific follow-up questions the system asks to authenticate and serve a request (Table 3).

Table 1. Sample layout of the Personal Information Database (PID).

Full name	NID number	Birth date	Recorded response (text)	GMM model	Call time
Abror Salimov	32007914010096	20/07/1991	Hisob ochish uchun... (To open an account...)	Person1.gmm	13.01.2026, 06:41 PM
Zafar Odilov	32806884050115	28/06/1988	Kartadan pul yechish... (Withdraw money from the card...)	Person2.gmm	14.01.2026, 06:56 PM

Table 2. Examples of Generic Information Database (GID) entries.

User query	System solution
Karta PIN kodini unutdim? (Forgot your card PIN?)	Apply to the bank branch where the card was opened, presenting an identity document.
SMS-xabarnoma xizmatini ulash (Connect the SMS notification service)	SMS notification for a national-currency plastic card can be connected through an infokiosk.
Bank xizmatlaridan qaysi kunlarda foydalanish mumkin? (On what days can bank services be used?)	Services are available Monday to Friday, except public holidays.

Table 3. Structure of the Credential Information Database (CID).

User response	Domain	Predefined follow-up questions
I have an account with this bank; I want to know my latest balance.	Balance issue	1. What is your account number? 2. Enter your NID number. 3. ...
My card is lost. What can I do?	Card issue	1. State your card number. 2. Enter your NID number. 3. ...

Experiments and Results Analysis

WaveNet Model Accuracy

The character error rate (CER) served as the principal evaluation benchmark. CER measures the proportion of characters that an ASR system identifies incorrectly, computed as $CER = (S + D + I)/N$, where S, D, and I denote substitutions, deletions, and insertions, respectively, and N the number of reference characters; lower values indicate better performance, with zero representing a perfect result. After a 24-hour training run on Colab Pro+ with an NVIDIA A100 GPU, the WaveNet model attained its best CER of 0.064907 at step 16,400, with final training and validation losses of 0.105120 and 0.332718 (Table 4).

Table 4. WaveNet model training progression.

Step	Training loss	Validation loss	CER
400	6.319100	3.322290	0.970519
800	2.128300	0.678408	0.207603
1200	0.879500	0.433805	0.141339
1600	0.736400	0.364600	0.121398
17600	0.108070	0.337852	0.065240
18000	0.106200	0.334034	0.065009
18400	0.105120	0.332718	0.064907

GPU-related constraints limited training to a 50-hour session on Colab Pro+. To characterize the effect of data volume, the training set was varied across 10,000, 15,000, and 25,000 sample

inputs (Table 5). The results confirm a strong correlation between the number of samples (and corresponding training time) and improvements in training loss, validation loss, and CER.

Table 5. Comparison among various sample-input sizes.

Sample inputs	Hours of training	Training loss	Validation loss	CER
10,000	20	1.830201	1.340452	0.840124
15,000	30	0.880291	0.643850	0.306281
25,000	50	0.105120	0.332718	0.064907

Seq2Seq-Based Summary Prediction

Sample data were drawn from the Uzbek text-summarization dataset. The word-count distribution shows that articles peaked at approximately 60 words, whereas reference summaries typically ranged between 5 and 10 words. During Seq2Seq training the step loss was 3.3613 and the validation loss 2.9232; the evolution of training and testing loss is shown in Figure 5.

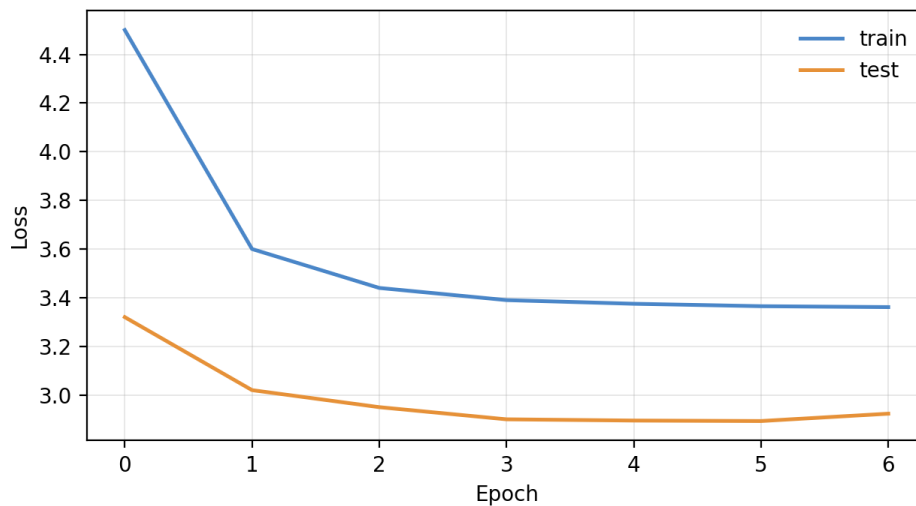


Figure 5. Training and testing value loss of the Seq2Seq summarization model.

The speaker-recognition system was tested in two settings-one with background noise and one with studio-level sound quality-using 56 samples across eight sequential stages per setting, monitoring accuracy as the sample count grew. In both scenarios the system identified all seven individuals when the sample count was limited to seven. When the sample size increased to 14, one individual was misrecognized in the noisy setting, lowering accuracy slightly to 96.4% (Figure 6). Although 96.4% accuracy was recorded under realistic conditions, effectiveness may diminish in larger-scale real-time environments.

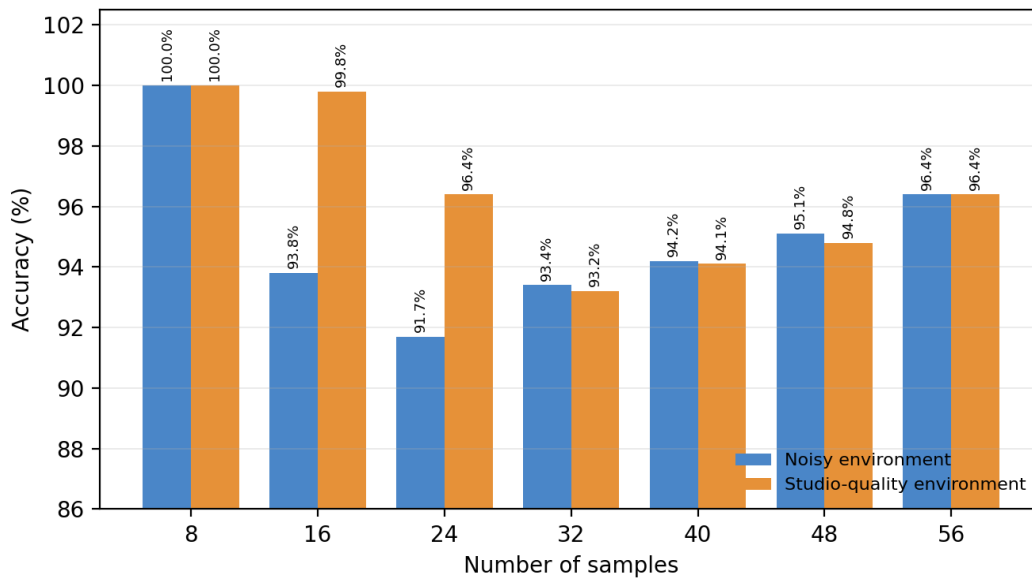


Figure 6. Accuracy of the speaker-recognition module in noisy versus studio-quality environments as a function of sample count (8-56 samples).

Sentence Similarity and Response Accuracy

The cosine-similarity algorithm quantified the resemblance between each customer inquiry and the questions in the existing dataset, and the most similar question-answer pairs were returned; the complete matching process takes approximately one to two minutes. Because customers phrase requests in many ways, the system was tested with varied question formats. The sentence-summarization model was trained with an RNN size of 256, batch size of 2, learning rate of 0.001, and dropout probability of 0.65-0.75 with the Adam optimizer, achieving an average loss of 0.004 on the Uzbek dataset. For sentence embedding, the BERT sentence transformer paraphrase-mpnet-base-v2 was fine-tuned and evaluated on four Uzbek datasets-our USC dataset, CC100-uzbek, voice-recognition-Uzbek, and xls-r-uzbek-cv8-achieving an average accuracy of 85.27% (Table 6).

Table 6. Accuracy of the paraphrase-mpnet-base-v2 sentence transformer in embedding paragraphs for search queries.

Dataset	Accuracy (%)
Our USC dataset	91.29
CC100-uzbek	88.96
voice-recognition-Uzbek	78.62
xls-r-uzbek-cv8	82.24

Tables 7 and 8 illustrate representative question variations and the corresponding system responses for two sample banking queries (credit services and card issuance). The system answered most paraphrases correctly-for example, correctly listing available loan types and microloan procedures-while occasional variants (e.g., an ambiguous “want to know about your credit?” phrasing, or a card-blocking question answered with unblocking instructions) produced incorrect responses.

Table 7. Question variations and system responses for the sample query “Siz kredit berasizmi?” (Do you give credit?).

Variation	System response (summary)	Assessment
What types of loans do you have?	Auto loans, microloans, consumer loans, and mortgage loans are available from banks.	Correct
How can I formalize a microloan?	Contact the credit manager, submit a loan application, and provide all documents required by the bank.	Correct
Want to know about your credit?	If you have an account number in our bank, you can apply for opening a debit card.	Incorrect
Can I pay the loan without going to the bank?	Yes-through the authenticated application, customers can make and monitor loan payments.	Correct

Table 8. Question variations and system responses for the sample query “Karta ochishim uchun nima qilishim kerak?” (What do I need to do to open a card?).

Variation	System response (summary)	Assessment
What is the procedure for enabling SMS notifications on the card?	Uzcard SMS notification can be activated via ATM; Humo SMS notification is enabled through the branch where the card was opened.	Correct
What should I do if I forget my card PIN?	Contact the branch where the card was opened, presenting an identity document.	Correct
How to block the card when it is lost?	To unblock the card, contact the branch where the card was opened with an identity document.	Incorrect
Is it possible to order a card online without visiting a branch?	All cards in national and foreign currency can be ordered remotely using the National application.	Correct

To assess overall precision, ten questions were randomly selected and at least four paraphrased variants were constructed for each (Figure 7). The system performed exceptionally well on most queries and produced appropriate responses, though some variants were answered inaccurately. On this evaluation the estimated end-to-end system accuracy is 82.5%, and we expect that enlarging the pool of variant questions will further improve accuracy.

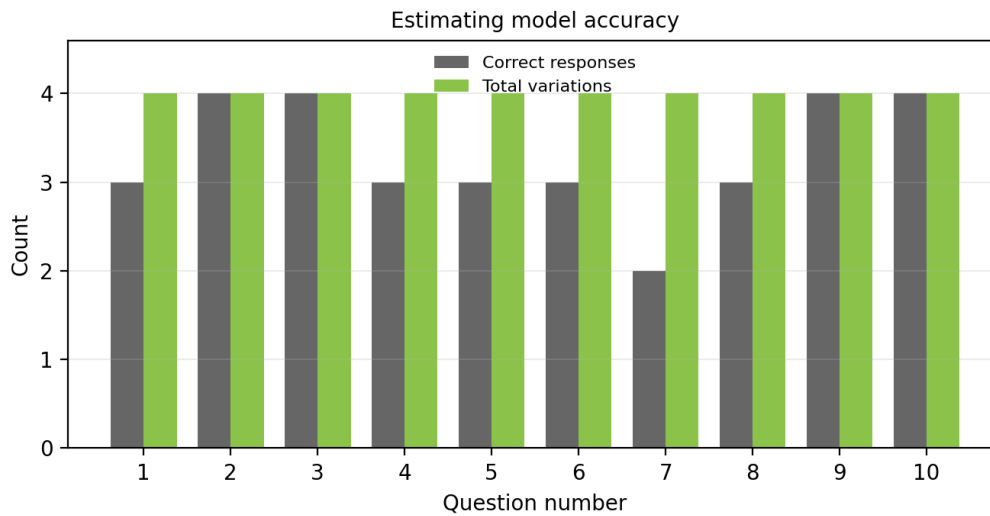


Figure 7. Estimated model accuracy: correct responses versus total variations for ten randomly selected questions.

ASR Model Comparison with Data Augmentation

To expand the corpus artificially, several augmentation techniques were applied: additive Gaussian noise (AN) with amplitude $\sigma = 0.01$; speed perturbation (SP) at 0.9, 1.0, and $1.1 \times$ playback rates; and spectral masking (SA) along the time and frequency axes, using two masks of maximum width $T = 40$ on the time axis and two masks of maximum width $F = 30$ on the frequency axis. The contribution of a language model (LM) at the decoding stage was also evaluated. Table 9 reports validation and test CER/WER for four architectures-E2E-BLSTM, DNN-CTC, E2E-Conformer, and Deep Speech 2-under incremental combinations of these techniques.

Table 9. Experimental results of model training with incremental augmentation (LM, AN, SP, SA). Values are CER/WER (%).

Model	Configuration	Valid CER	Valid WER	Test CER / WER
E2E-BLSTM	Baseline	13.9	43.2	15.1 / 44.7
	+ LM + AN + SP + SA	10.6	22.6	11.5 / 23.9
DNN-CTC	Baseline	13.1	35.1	10.9 / 32.7
	+ LM + AN + SP + SA	6.0	17.1	6.5 / 21.9
E2E-Conformer	Baseline	9.3	40.5	12.6 / 44.2
	+ LM + AN + SP + SA	7.6	23.1	8.9 / 22.3
Deep Speech 2	Baseline	12.0	36.7	10.2 / 34.6
	+ LM + AN + SP + SA	5.4	15.1	5.22 / 13.8

Across the experiments, Deep Speech 2 demonstrated superior performance in both WER and CER. On a test set of 5,474 samples, the fully augmented Deep Speech 2 model achieved a WER of 13.8% and a CER of 5.22%, confirming its effectiveness in accurately processing and transcribing Uzbek speech. The results also show that each augmentation technique contributes

incrementally, with the combination of language-model decoding, noise addition, speed perturbation, and spectral masking yielding the largest gains for every architecture.

5. Conclusions and Future Work

UzAssistant represents a significant step toward intelligent service automation in Uzbekistan's banking sector. By recognizing callers, transcribing and understanding Uzbek speech, retrieving relevant answers, and responding with natural synthesized speech, the system offers a transformative approach to how banking services are accessed: it can elevate the customer experience, broaden financial inclusion, and reduce operational expenses by providing round-the-clock client service without additional staff. A key benefit is the ability to interact with customers in their native Uzbek language, which greatly enhances user satisfaction in a market historically served only by English- and Russian-language speech technology.

At the same time, automated voice-chat systems face challenges that can affect the user experience, including limited vocabulary coverage, residual speech-recognition errors, constrained multi-modal interaction, a somewhat impersonal tone, and potential technical faults. Despite these challenges, the potential for advancement in the Uzbek market is substantial. Promising research directions include richer multi-modal interaction; more advanced learning methods trained on larger datasets; fully real-time response generation; additional functionality such as handling complex transactions and personalized recommendations based on customers' banking histories; and built-in feedback mechanisms for gathering client opinions on precision and effectiveness. Recent progress in TTS technology already makes it possible to produce increasingly lifelike automated voices in Uzbek; pursuing these directions will yield a more immersive and effective customer experience, ultimately benefiting both customers and banking institutions.

Reference

1. Guzman, A.L. Voices in and of the machine: Source orientation toward mobile virtual assistants. *Comput. Hum. Behav.* 2018, 90, 343-350.
2. McCue, T.J. Okay Google: Voice search technology and the rise of voice commerce. *Forbes Online*, 2018.
3. Juniper Research. Voice assistants used in smart homes to grow 1000%, reaching 275 million by 2023, as Alexa leads the way. 2018.
4. Gartner. Digital assistants will serve as the primary interface to the connected home. *Gartner Online*, 2016.
5. Hoy, M.B. Alexa, Siri, Cortana, and more: An introduction to voice assistants. *Med. Ref. Serv. Q.* 2018, 37, 81-88.
6. Sergey, O. Listens and understands: How automatic speech recognition technology works. Available online (accessed 4 March 2023).
7. Mukhamadiyev, A.; Khujayarov, I.; Djuraev, O.; Cho, J. Automatic speech recognition method based on deep learning approaches for Uzbek language. *Sensors* 2022, 22, 3683.
8. Mukhamadiyev, A.; Mukhiddinov, M.; Khujayarov, I.; Ochilov, M.; Cho, J. Development of language models for continuous Uzbek speech recognition system. *Sensors* 2023, 23, 1145.

9. Ochilov, M. Social network services-based approach to speech corpus creation. TUIT News 2021, 1(57), 21-31.
10. Musaev, M.; Mussakhojayeva, S.; Khujayorov, I.; Khassanov, Y.; Ochilov, M.; Varol, H.A. USC: An open-source Uzbek speech corpus and initial speech recognition experiments. In Proceedings of SPECOM 2021, St. Petersburg, Russia, 27-30 September 2021.
11. Khujayorov, I.Sh.; Ochilov, M.M. Analysis of methods of acoustic modeling of speech signals based on neural networks. TUIT News 2020, 2(54), 2-15.
12. Musaev, M.; Khujayorov, I.; Ochilov, M. Image approach to speech recognition on CNN. In Proceedings of ICFNN 2019, Rome, Italy, 26-28 July 2019; pp. 1-6.
13. Sundar, S.S.; Jung, E.H.; Waddell, F.T.; Kim, K.J. Cheery companions or serious assistants? Role and demeanour congruity as predictors of robot attraction and use intentions among senior citizens. *Int. J. Hum.-Comput. Stud.* 2017, 97, 88-97.
14. Balakrishnan, J.; Dwivedi, Y.K. Conversational commerce: Entering the next stage of AI-powered digital assistants. *Ann. Oper. Res.* 2021, 1-35.
15. Liao, Y.; Vitak, J.; Kumar, P.; Zimmer, M.; Kritikos, K. Understanding the role of privacy and trust in intelligent personal assistant adoption. In Proceedings of the International Conference on Information, 2019.
16. Moriuchi, E. Okay, Google!: An empirical study on voice assistants on consumer engagement and loyalty. *Psychol. Mark.* 2019, 36, 489-501.
17. McLean, G.; Osei-Frimpong, K. Hey Alexa... examine the variables influencing the use of artificial intelligent in-home voice assistants. *Comput. Hum. Behav.* 2019, 99, 28-37.
18. Pantano, E.; Pizzi, G. Forecasting artificial intelligence on online customer assistance: Evidence from chatbot patents analysis. *J. Retail. Consum. Serv.* 2020, 55, 102096.
19. Smith, S. Voice assistants used in smart homes to grow 1,000%, reaching 275 million by 2023, as Alexa leads the way. Juniper Research, 2018.
20. Goasduff, L. Chatbots will appeal to modern workers. Gartner, 2019.
21. Swoboda, C. COVID-19 is making Alexa and Siri a hands-free necessity. *Forbes*, 2020.
22. Barnes, S.J. Information management research and practice in the post-COVID-19 world. *Int. J. Inf. Manag.* 2020, 55, 102175.
23. Carroll, N.; Conboy, K. Normalising the “new normal”: Changing tech-driven work practices under pandemic time pressure. *Int. J. Inf. Manag.* 2020, 55, 102186.
24. Papagiannidis, S.; Harris, J.; Morton, D. WHO led the digital transformation of your company? A reflection of IT related challenges during the pandemic. *Int. J. Inf. Manag.* 2020, 55, 102166.
25. Marikyan, D.; Papagiannidis, S.; Alamanos, E. A systematic review of the smart home literature: A user perspective. *Technol. Forecast. Soc. Change* 2019, 138, 139-154.
26. Balakrishnan, J.; Dwivedi, Y.K. Conversational commerce: Entering the next stage of AI-powered digital assistants. *Ann. Oper. Res.* 2021, 1-35.
27. Benbya, H.; Davenport, T.H.; Pachidi, S. Artificial intelligence in organizations: Current state and future opportunities. *MIS Q. Exec.* 2020, 19(4).

28. Fernandes, T.; Oliveira, E. Understanding consumers' acceptance of automated technologies in service encounters: Drivers of digital voice assistants adoption. *J. Bus. Res.* 2021, 122, 180-191.
29. Hamet, P.; Tremblay, J. Artificial intelligence in medicine. *Metabolism* 2017, 69, S36-S40.
30. Li, B.-H.; Hou, B.-C.; Yu, W.-T.; Lu, X.-B.; Yang, C.-W. Applications of artificial intelligence in intelligent manufacturing: A review. *Front. Inf. Technol. Electron. Eng.* 2017, 18, 86-96.
31. Olshannikova, E.; Ometov, A.; Koucheryavy, Y.; Olsson, T. Visualizing Big Data with augmented and virtual reality: Challenges and research agenda. *J. Big Data* 2015, 2, 1-27.
32. Young, A.G.; Majchrzak, A.; Kane, G.C. Organizing workers and machine learning tools for a less oppressive workplace. *Int. J. Inf. Manag.* 2021, 59, 102353.
33. Kane, G.C.; Young, A.G.; Majchrzak, A.; Ransbotham, S. Avoiding an oppressive future of machine learning: A design theory for emancipatory assistants. *MIS Q.* 2021, 45, 371-396.
34. Schwenk, H.; Gauvain, J.L. Training neural network language models on very large corpora. In *Proceedings of HLT/EMNLP*, Vancouver, BC, Canada, October 2005; pp. 201-208.
35. Mikolov, T.; Karafiát, M.; Burget, L.; Cernocký, J.; Khudanpur, S. Recurrent neural network based language model. In *Proceedings of Interspeech 2010*; Vol. 3, pp. 1045-1048.
36. Huang, Z.; Zweig, G.; Dumoulin, B. Cache based recurrent neural network language model inference for first pass speech recognition. In *Proceedings of ICASSP 2014*; IEEE: Piscataway, NJ, USA, 2014; pp. 6354-6358.
37. Sundermeyer, M.; Oparin, I.; Gauvain, J.L.; Freiberg, B.; Schlüter, R.; Ney, H. Comparison of feedforward and recurrent neural network language models. In *Proceedings of ICASSP 2013*; IEEE: Piscataway, NJ, USA, 2013; pp. 8430-8434.
38. Morioka, T.; Iwata, T.; Hori, T.; Kobayashi, T. Multiscale recurrent neural network based language model. In *Proceedings of Interspeech 2015*, Dublin, Ireland.
39. Hardy, H.; Strzalkowski, T.; Wu, M. Dialogue Management for an Automated Multilingual Call Center; State University of New York at Albany, Institute for Informatics, Logics and Security Studies: Albany, NY, USA, 2003.
40. Zweig, G.; et al. Automated quality monitoring for call centers using speech and NLP technologies. In *Proceedings of the HLT-NAACL Companion Volume: Demonstrations*, 2006.
41. McLean, G.; Osei-Frimpong, K. Examining satisfaction with the experience during a live chat service encounter-implications for website providers. *Comput. Hum. Behav.* 2017, 76, 494-508.
42. Warnapura, A.K.; et al. Automated customer care service system for finance companies. Research and Publication of Sri Lanka Institute of Information Technology (SLIIT), NCTM, 2014.
43. Mansurov, B.; Mansurov, A. UzBERT: Pretraining a BERT model for Uzbek. *arXiv* 2021, arXiv:2108.09814.

44. Ren, Z.; Yolwas, N.; Slamu, W.; Cao, R.; Wang, H. Improving hybrid CTC/attention architecture for agglutinative language speech recognition. *Sensors* 2022, 22, 7319.
45. Mamatov, N.S.; Niyozmatova, N.A.; Abdullaev, S.S.; Samijonov, A.N.; Erejepov, K.K. Speech recognition based on transformer neural networks. In *Proceedings of ICISCT 2021*; IEEE: Piscataway, NJ, USA, 2021; pp. 1-5.
46. Le, Q.; Mikolov, T. Distributed representations of sentences and documents. In *Proceedings of ICML 2014*; PMLR, 2014; pp. 1188-1196.
47. Mikolov, T.; Chen, K.; Corrado, G.; Dean, J. Efficient estimation of word representations in vector space. *arXiv* 2013, arXiv:1301.3781.
48. Khamdamov, U.; Mukhiddinov, M.; Akmuradov, B.; Zarmasov, E. A novel algorithm of numbers to text conversion for Uzbek language TTS synthesizer. In *Proceedings of ICISCT 2020*, Tashkent, Uzbekistan; pp. 1-5.
49. Zhao, Q.; Tu, D.; Xu, S.; Shao, H.; Meng, Q. Natural human-robot interaction for elderly and disabled healthcare application. In *Proceedings of IEEE BIBM 2014*; pp. 39-44.
50. Yan, H.; Ang, M.H.; Poo, A.N. A survey on perception methods for human-robot interaction in social robots. *Int. J. Soc. Robot.* 2014, 6, 85-119.
51. Amodei, D.; et al. Deep Speech 2: End-to-end speech recognition in English and Mandarin. In *Proceedings of ICML 2016*; PMLR, 2016; pp. 173-182.
52. Sultana, M.; Chakraborty, P.; Choudhury, T. Bengali abstractive news summarization using Seq2Seq learning with attention. In *Cyber Intelligence and Information Retrieval*; Springer: Singapore, 2022; pp. 279-289.
53. van den Oord, A.; Dieleman, S.; Zen, H.; Simonyan, K.; Vinyals, O.; Graves, A.; Kalchbrenner, N.; Senior, A.; Kavukcuoglu, K. WaveNet: A generative model for raw audio. *arXiv* 2016, arXiv:1609.03499.
54. Okamoto, T.; Toda, T.; Shiga, Y.; Kawai, H. Tacotron-based acoustic model using phoneme alignment for practical neural text-to-speech systems. In *Proceedings of ASRU 2019*; pp. 214-221.
55. Bishop, C.M. *Mixture Density Networks*; Technical Report; 1994.
56. Quatieri, T.F. *Discrete-Time Speech Signal Processing: Principles and Practice*; Pearson Education India, 2006.
57. Tamamori, A.; Hayashi, T.; Kobayashi, K.; Takeda, K.; Toda, T. Speaker-dependent WaveNet vocoder. In *Proceedings of Interspeech 2017*; pp. 1118-1122.
58. van den Oord, A.; et al. WaveNet: A generative model for raw audio. *arXiv* 2016. Available online: <https://arxiv.org/pdf/1609.03499.pdf>.
59. Song, E.; Soong, F.K.; Kang, H.-G. Effective spectral and excitation modeling techniques for LSTM-RNN-based speech synthesis systems. *IEEE/ACM Trans. Audio Speech Lang. Process.* 2017, 25, 2152-2161.
60. Salimans, T.; Kingma, D.P. Weight normalization: A simple reparameterization to accelerate training of deep neural networks. In *Proceedings of NIPS 2016*; pp. 901-909.

Governance-Embedded Scrum for AIoT in Safety-Critical Cyber-Physical Systems: A Maritime Smart-Ship Case

N.S. Saydirasulov

Department of Computer Engineering, Gachon University, Seongnam-si, Republic of Korea;
email: saydirasulov@gachon.ac.kr

D.A. Davronbekov

Department of Technologies of Mobile Communication, Tashkent University of Information Technologies Named After Muhammad Al-Khwarizmi, Tashkent 100200, Uzbekistan,
email: d.davronbekov@gmail.com

Young Im Cho

Department of Computer Engineering, Gachon University, Seongnam-si, Republic of Korea

Abstract: Smart ships are increasingly built as AIoT-enabled cyber-physical systems, yet safety-critical maritime software must preserve deterministic control, auditability and certification evidence while still allowing rapid service evolution. This article presents a shortened journal version of a governance-embedded Scrum framework that couples a certifiable monolithic safety core with modular AIoT services through a secure API gateway. The gateway enforces authentication and authorization, schema and contract conformance, rate and priority control, and auditable mediation between read-only service views and protected safety buses. Governance checkpoints are embedded into sprint planning, daily monitoring, sprint review and retrospective activities, so that security, compliance, resilience and traceability evidence are produced within each increment rather than added after development. Validation was conducted through a digital-twin simulation, a hardware-in-the-loop laboratory rig and an integrated prototype emulating more than 500 shipboard devices under nominal, adverse, security and update scenarios. Compared with cloud-centric microservices and a hybrid baseline without governance, the proposed framework reduced p95 latency by up to 38%, shortened mean time to recovery by up to 55%, improved deployment lead time by 28%, and maintained approximately 99.9% reliability in controlled tests. The results indicate that governed interfaces and in-sprint assurance can reduce the safety-innovation gap in maritime AIoT, although live-vessel validation remains future work.

Keywords: AIoT; Scrum governance; hybrid architecture; maritime safety; cyber-physical systems; secure API gateway; reliability; smart ships

Introduction

Maritime digitalization is moving from isolated shipboard automation toward AIoT-enabled smart ships in which sensors, embedded controllers, decision-support tools and networked services operate as a single cyber-physical system. Predictive maintenance, route and energy optimization, environmental monitoring and crew-facing decision support can improve safety and efficiency, but the deployment context is unusually demanding. Navigation, propulsion and alarm functions must

remain deterministic and auditable even when connectivity is degraded, workloads are bursty or software services are updated during operation [1-5].

The core challenge is the safety-innovation paradox. Monolithic safety architectures provide strong timing guarantees, fault containment and evidence for certification, yet they slow the introduction of new AI services. Conversely, modular and cloud-centric microservice architectures increase development velocity, but can introduce long-tail latency, interoperability fragility and weaker isolation under link impairment or adversarial traffic [6-9]. Maritime AIoT therefore requires an architecture and a delivery process that allow innovation without exposing safety-critical paths to uncontrolled change.

This study proposes a governance-embedded Scrum framework for AIoT in safety-critical maritime CPS. The framework combines a monolithic safety core, a secure API gateway and modular AIoT services. The safety core handles typed device abstraction, deterministic control and audit logs; the gateway mediates contracts, authorization, rate limits and priority classes; and the service layer hosts predictive maintenance, energy optimization and decision-support modules. Scrum is extended with in-sprint governance gates, meaning that cybersecurity, compliance, resilience and traceability evidence are treated as backlog work and release criteria rather than separate post-hoc audits [10-13].

The contribution of this shortened journal manuscript is threefold: (i) a compact hybrid architecture that separates certifiable control from modular AIoT innovation; (ii) an operational Scrum extension that binds governance evidence to each sprint increment; and (iii) a three-stage evaluation using digital-twin simulation, HIL laboratory testing and an integrated prototype with 500+ emulated shipboard devices. The article is intentionally condensed to suit journal submission and excludes extended pseudocode and large scenario appendices from the original long version.

Related Work and Research Gap

AIoT applications in maritime systems have expanded rapidly in predictive maintenance, voyage planning, energy management and situational awareness. Recent studies show that multi-sensor models can detect machinery anomalies and support maintenance planning, while route and energy optimization can reduce fuel use under controlled assumptions [3,4,14-17]. However, many AI-centric deployments rely on bespoke middleware or cloud backhaul and do not explicitly model timing isolation, write protection or audit evidence as first-class design variables.

Architectural research presents a tension between monolithic, microservice and hybrid patterns. Monolithic safety cores fit deterministic control and certification practices, but create high coupling and slow release cycles. Microservices improve scalability and independent deployment, but introduce extra communication paths and operational complexity. Hybrid edge-first designs are therefore attractive for maritime CPS, provided that interfaces, safety boundaries and release conditions are explicitly governed [6-9,18].

Agile methods have also been adapted for regulated domains. SafeScrum and related approaches show that iterative development can coexist with assurance activities when safety requirements, traceability and verification are represented in the backlog [10-13]. Nevertheless, maritime AIoT imposes continuous compliance during voyages, mixed-criticality workloads, vendor heterogeneity and cyber-risk management requirements that are often absent from general Agile case studies. The gap addressed here is a repeatable, audit-ready pattern that integrates governed hybrid interfaces with Scrum ceremonies and validates them under realistic maritime impairments.

Table 1. Condensed positioning of the proposed framework.

Approach	Strength	Main weakness	Governance level	Fit for maritime AIoT
Monolithic safety core	Determinism, auditability	Low service velocity	External or post-hoc	Strong safety, weak innovation
Cloud-centric microservices	Fast service evolution	Latency and isolation risk	Often operational only	Weak for safety-critical loops
Hybrid without gates	Core protection plus modularity	Regression risk across sprints	Partial checklists	Useful but not audit-ready
Proposed hybrid + Scrum governance	Bounded interfaces and continuous evidence	Needs organizational discipline	Embedded in sprint cadence	Balanced safety and innovation

Materials and Methods

Hybrid Maritime AIoT Architecture

The proposed architecture separates safety-critical functions from innovation-oriented services. The monolithic safety core provides typed hardware abstraction, health monitoring, deterministic control loops and immutable audit logging. It remains the only component allowed to write to protected safety buses. Modular AIoT services consume read-only, normalized data views and return recommendations or analytics through governed interfaces rather than direct actuator commands.

The secure API gateway is the architectural policy-enforcement point. It verifies TLS 1.3 and mutual authentication, validates OAuth 2.0/OIDC tokens and scoped access, checks JSON-Schema or Protobuf contracts, applies semantic validation against maritime data models, and assigns rate and priority classes. Requests that would violate authorization, schemas, contracts or SLO budgets are rejected, queued or shed before reaching the safety core. This design makes the gateway a practical safety boundary as well as an interoperability layer.

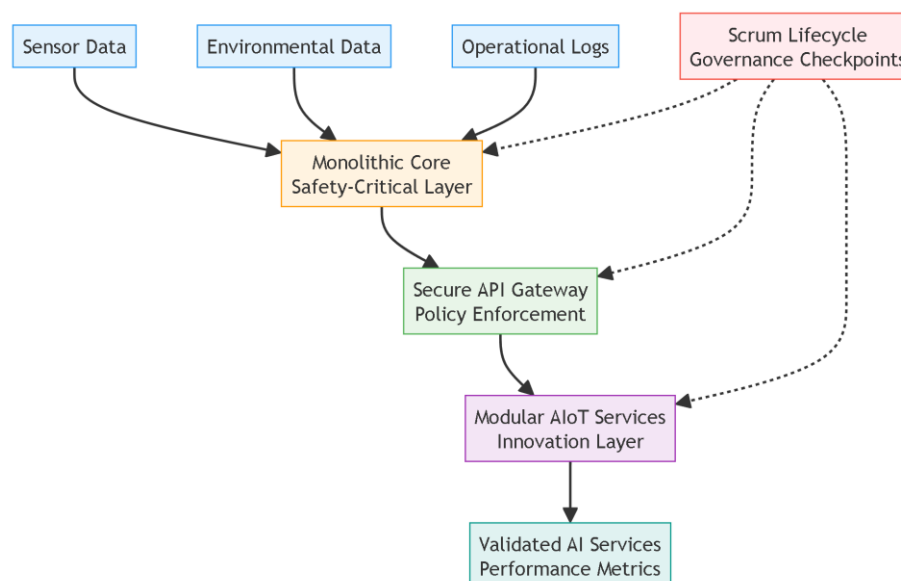


Figure 1. Proposed smart-ship AIoT framework: sensor, environmental and operational data are processed by a monolithic safety core; a secure API gateway mediates modular services and the Scrum lifecycle provides governance checkpoints.

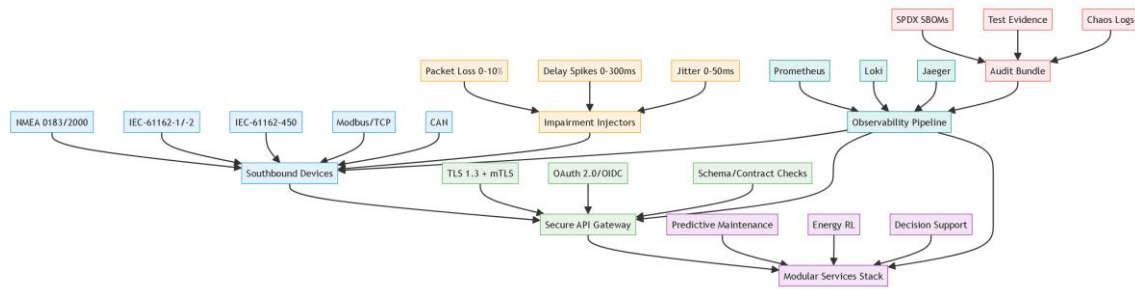


Figure 2. Digital-twin validation topology showing southbound maritime protocols, impairment injection, gateway mediation and observability artifacts used by the release gate.

Governance-Embedded Scrum Process

Classical Scrum is extended so that assurance work is part of the sprint, not a separate certification phase. During backlog refinement and sprint planning, feature stories are paired with compliance tasks, threat-model updates, test-evidence tasks, SBOM generation and resilience experiments. Daily standups track both implementation progress and evidence burndown. Sprint review functions as a release gate in which security scans, contract tests, SLO checks, traceability evidence and rollback readiness must pass before deployment.

Retrospectives convert near-misses, nuisance alarms, failed policy checks and operator feedback into new guardrails. This closes the loop between runtime telemetry and backlog refinement. The Definition of Done therefore includes not only functional correctness but also passed SAST/DAST/SCA checks, contract conformance, WCET margin evidence, chaos-test results, audit logs and signed release attestations.

Validation Design

Evaluation was organized in three stages. First, a digital twin emulated 500-1,200 shipboard endpoints spanning navigation, propulsion, environmental and hotel-load subsystems. Traffic included 1-10 Hz telemetry, alarm bursts up to 200 messages per second and bulk log uploads. Network impairment profiles included packet loss up to 10%, delay spikes up to 300 ms, jitter up to 50 ms, link flaps, sensor freezes and outliers.

Second, a hardware-in-the-loop laboratory rig used embedded controllers, CAN transceivers, industrial I/O, RS-422 and Ethernet links to test timing determinism and write protection on real I/O chains. Third, an integrated prototype evaluated CI/CD with shadow, canary and blue-green releases under automatic rollback. The methods compared four variants: monolithic-only, cloud-centric microservices, hybrid without governance and the proposed hybrid with governance.

Table 2. Validation scenarios and primary metrics.

Scenario	Workload / event	Key metrics	Acceptance focus
S-A Nominal	Periodic telemetry, alarm bursts, calm-to-rough sea profiles	p50/p95/p99 latency, jitter	Sub-100 ms p95 on critical paths
S-B Adverse	Loss, delay, jitter, link flaps, sensor freezes	Availability, MTTR, rollback time	No cascading failures
S-C Security	Token expiry, replay, malformed payloads, contract violations	Policy violations, blocked writes	Zero unauthorized safety-bus writes
S-D Updates	Shadow, canary and blue-green releases	Deployment lead time, audit pass rate	Release only after governance gate

Governance Release Checklist and Reproducibility

To make the framework implementable, the sprint release gate is expressed as a checklist that can be automated in CI/CD. Each candidate increment must produce evidence for security, compliance, resilience and operational performance. Security evidence includes static and dynamic analysis, software composition analysis, SBOM generation and credential hygiene checks. Compliance evidence links requirements to tests and confirms that safety-related contracts remain unchanged or have been reviewed.

Resilience evidence is produced through chaos and impairment experiments. The release candidate is exposed to packet loss, delay, jitter, link flap and malformed-input scenarios before promotion. If a service violates the SLO budget or triggers unsafe contract behavior during canary exposure, the deployment is automatically rolled back and the failed evidence is returned to the backlog. This design turns operational failure modes into sprint-level learning items.

Reproducibility is supported by fixed random seeds, reusable workload manifests, containerized services and standardized telemetry exports. The same scenario definitions are applied to all baselines, and each run stores logs, traces, metrics and release-gate outcomes. Although proprietary ship data cannot always be shared, sanitized synthetic generators and configuration files can allow other researchers to reproduce timing and governance behavior.

The checklist also supports journal reporting because it clarifies which results come from simulation, which from HIL and which from the integrated prototype. This separation prevents overstating the claims: the study demonstrates controlled feasibility and measurable improvement, not formal vessel certification.

Table 3. Sprint release-gate checklist for the proposed framework.

Gate	Evidence artifact	Automated check	Release action
Security	SAST, DAST, SCA, SBOM	No critical findings or unmanaged CVEs	Approve or block
Compliance	Requirements-to-test matrix, contract diff	Traceability complete; schema reviewed	Approve or require review
Resilience	Chaos logs, impairment results	SLO preserved or rollback succeeds	Canary or reject
Performance	Latency, jitter, throughput dashboard	p95 target and WCET margin maintained	Promote or tune

Operations	Runbook and rollback plan	Telemetry and alert routing available	Blue-green release
------------	---------------------------	---------------------------------------	--------------------

Results

The proposed framework improved both operational performance and assurance outcomes. Across nominal and adverse operation, p95 latency was reduced by 38% relative to the cloud-centric microservice baseline and by 23% relative to the hybrid baseline without governance. p99 latency also improved, indicating that the gateway reduced long-tail effects by rejecting malformed, unauthorized or out-of-budget requests early.

Availability remained approximately 99.9% in the controlled simulation and prototype runs. Under adverse network conditions, automatic rollback, circuit breaking and controlled shedding of non-critical traffic reduced mean time to recovery by up to 55%. These gains were strongest in S-B and S-D, where network impairment and update campaigns created the highest risk of cascading errors.

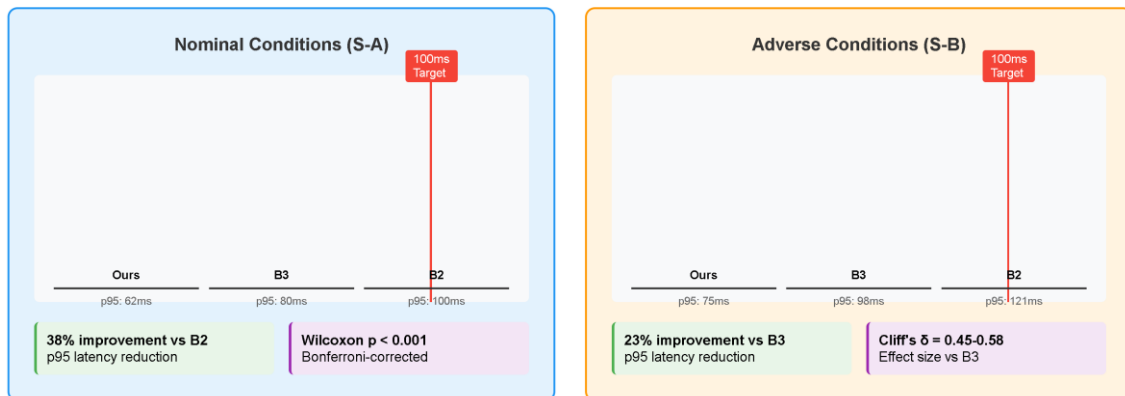


Figure 3. Condensed latency result: the proposed approach shifts the tail of the latency distribution left under nominal and adverse conditions, preserving the sub-100 ms target more consistently than baselines.

Table 4. Key performance and assurance results.

KPI	Baseline comparison	Observed result	Interpretation	Evidence stage
p95 latency	Cloud-centric services	-38%	Lower long-tail delay	Digital twin + prototype
MTTR	Hybrid without governance	-55%	Faster rollback and recovery	Adverse scenarios
Reliability / availability	All baselines under impairment	$\approx 99.9\%$	SLO preservation	Simulation + prototype
Safety-bus writes	Adversarial attempts	0 successful writes	Isolation preserved	HIL rig
Deployment lead time	Ungoverned release pipeline	-28%	Faster certified increments	CI/CD prototype
Fuel-comfort proxy	Baseline policy	+12 pp	Energy-aware human-centric trade-off	Simulation

Security outcomes confirmed the importance of runtime policy enforcement. In HIL testing, no adversarial test obtained write access to the protected safety bus. Policy-violation rates were

reduced by roughly an order of magnitude compared with the cloud-centric and ungated hybrid variants. This occurred because authentication, scoped authorization, schema validation and contract validation were performed before requests reached the core.

Ablation results indicate that the effect is not caused by a single AI service. Removing governance gates increased p99 latency and MTTR and lowered audit pass rate; disabling gateway policy checks had an even stronger negative effect on security metrics. Isolating predictive-maintenance-only or reinforcement-learning-only services produced smaller changes, suggesting that architecture and process controls account for most of the observed improvement.

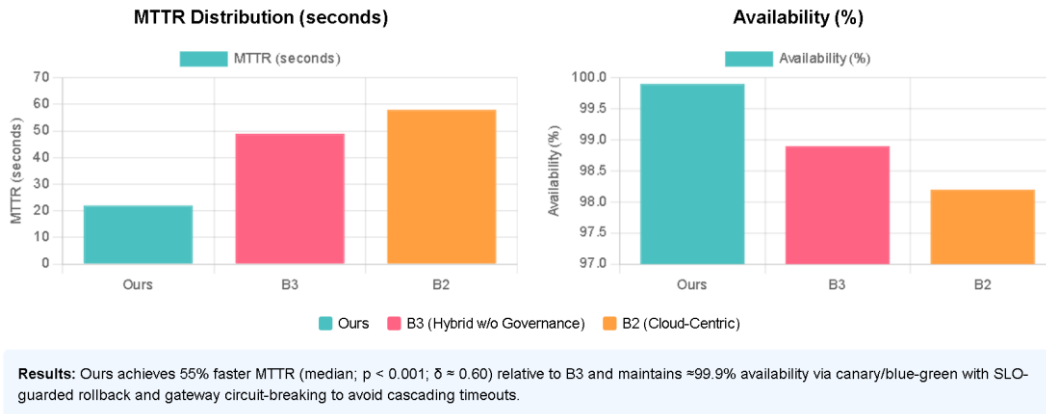


Figure 4. Repair time and availability under adverse conditions: SLO-guarded rollback and gateway circuit breaking contract MTTR while maintaining higher availability.

Discussion

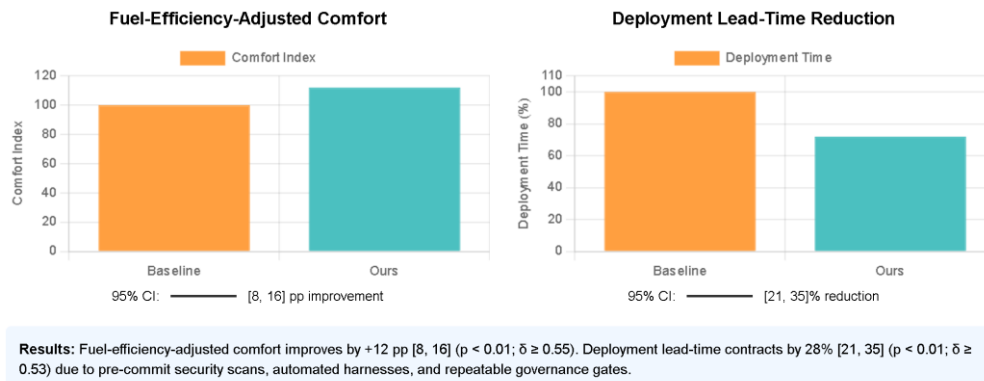


Figure 5. Industry 5.0-oriented outcomes: the comfort proxy improved while deployment lead time decreased, indicating that assurance automation need not slow service delivery.

The results support the argument that maritime AIoT cannot be governed by architecture alone or by process alone. The secure gateway creates an enforceable runtime boundary, but governance gates prevent unsafe or performance-regressing builds from reaching that boundary. Conversely, Scrum governance produces evidence and discipline, but without runtime contracts the safety core would still be exposed to malformed requests, retry storms or unauthorized calls. The proposed framework is therefore a socio-technical pattern: architectural contracts are translated into sprint tasks and sprint gates are enforced through runtime mechanisms.

The practical implication is that AIoT service teams can innovate at the periphery while the safety kernel remains stable. Predictive maintenance models, decision support and energy optimization can be retrained and redeployed through canary exposure, while the gateway ensures read-only service views and protects deterministic control loops. This is particularly important for

mixed-criticality vessels where bridge, engine, hotel-load and analytics functions share infrastructure but have different safety budgets.

The study also shows that governance can increase delivery speed when it is automated and repeatable. Although additional checks appear to add process overhead, pre-commit scanning, reusable test harnesses, SBOM generation and standardized release gates reduced rework and shortened deployment lead time in the prototype. In this sense, governance is not a late bureaucratic layer; it becomes a mechanism for reducing uncertainty and improving release confidence.

Limitations must be stated clearly. The evaluation used digital-twin, HIL and prototype testbeds; it did not include live-vessel trials. Real ships may introduce satellite backhaul variability, GNSS anomalies, crew workload peaks, vendor-specific data formats, classification-society constraints and operational practices that are only partially captured here. The reported reliability and latency values therefore demonstrate controlled feasibility rather than formal certification or operational approval.

From a submission perspective, the manuscript should be aligned with the target MDPI journal scope. For maritime communication and automation audiences, the most important framing is the safety-critical CPS contribution; for software engineering audiences, the governance process and release-evidence model should be emphasized. The present shortened version keeps both aspects visible, but a final submission can further adjust the emphasis, figure count and reference set according to the selected journal.

Conclusions

This article presented a compact MDPI-style journal manuscript on governance-embedded Scrum for AIoT in safety-critical maritime cyber-physical systems. The proposed framework combines a monolithic safety core, a secure API gateway and modular AIoT services, while embedding cybersecurity, compliance, resilience and traceability checkpoints into the sprint cadence. In controlled digital-twin, HIL and prototype evaluations, the approach reduced p95 latency, shortened MTTR, improved deployment lead time and maintained near-99.9% reliability while blocking unauthorized safety-bus writes.

Future work should validate the framework on operating vessels under classification-society oversight, extend formal verification of the safety kernel, and replicate the pattern in other safety-critical CPS domains such as aviation, healthcare and industrial automation. Additional work should also refine the human-centered layer, including explainable decision support, alarm fatigue management and crew workload metrics.

Reference

1. UNCTAD. Review of Maritime Transport 2025; United Nations: Geneva, Switzerland, 2025.
2. European Commission. Industry 5.0: Towards a Sustainable, Human-Centric and Resilient European Industry; Publications Office of the European Union: Luxembourg, 2021.
3. Shantharama, P.C.; Guddeti, R.M. A survey on Artificial Intelligence of Things: enabling technologies and applications. *Future Generation Computer Systems* 2022, 134, 189-210.
4. Dias, A.M.F.; et al. Digital twins in maritime systems: A survey. *Ocean Engineering* 2023, 279, 114500.
5. International Maritime Organization. Revised IMO Strategy on Reduction of GHG Emissions from Ships; IMO: London, UK, 2023.
6. Newman, S. Building Microservices, 2nd ed.; O'Reilly Media: Sebastopol, CA, USA, 2021.

7. Pahl, C.; Jamshidi, P. Microservices: A systematic mapping study. In Proceedings of CLOSER, Porto, Portugal, 2016; pp. 137-146.
8. Dragoni, N.; Giallorenzo, S.; Lafuente, A.L.; et al. Microservices: Yesterday, today, and tomorrow. In Present and Ulterior Software Engineering; Springer: Cham, Switzerland, 2017; pp. 195-216.
9. Chen, T.; Kim, M. Achieving latency SLOs in microservice architectures. In Proceedings of ICSA, Honolulu, HI, USA, 2022.
10. Schwaber, K.; Sutherland, J. The Scrum Guide: The Definitive Guide to Scrum; Scrum.org, 2020.
11. Stalhane, T.; Myklebust, J. SafeScrum: Development of safety-critical software using Scrum. In Proceedings of CISIM, Vilnius, Lithuania, 2016.
12. Dingsoyr, T.; Moe, N.B.; Tonelli, R. Agile in regulated environments: Challenges and evidence. *Empirical Software Engineering* 2023, 28, 1-30.
13. Leveson, N.G. Engineering a Safer World: Systems Thinking Applied to Safety; MIT Press: Cambridge, MA, USA, 2011.
14. Simon, H.A. The architecture of complexity. *Proceedings of the American Philosophical Society* 1962, 106, 467-482.
15. Baldwin, C.Y.; Clark, K.B. Design Rules: The Power of Modularity; MIT Press: Cambridge, MA, USA, 2000.
16. Liu, Y.; Chen, X. Digital twin for ships: Technologies and applications. *Ocean Engineering* 2022, 257, 111682.
17. Ahmed, S.; Hossain, M.S. Data-driven predictive maintenance for ship engines. *Ocean Engineering* 2023, 267, 113247.
18. Wang, H.; Sun, Z. Multi-objective route planning for ships under uncertain weather. *Applied Ocean Research* 2022, 120, 103050.
19. IEC. IEC 61508: Functional Safety of Electrical/Electronic/Programmable Electronic Safety-Related Systems; International Electrotechnical Commission: Geneva, Switzerland, 2010-2017.
20. IEC. IEC 61162-450: Maritime Navigation and Radiocommunication Equipment and Systems-Digital Interfaces-Ethernet Interconnection; IEC: Geneva, Switzerland, 2011.
21. IEC. IEC 62443-3-3: Industrial Communication Networks-Network and System Security-System Security Requirements and Security Levels; IEC: Geneva, Switzerland, 2013.
22. IETF. RFC 8446: The Transport Layer Security (TLS) Protocol Version 1.3; Internet Engineering Task Force, 2018.
23. IETF. RFC 6749: The OAuth 2.0 Authorization Framework; Internet Engineering Task Force, 2012.
24. OpenID Foundation. OpenID Connect Core 1.0 Incorporating Errata Set 2; OpenID Foundation, 2014.
25. NIST. SP 800-82 Rev. 3: Guide to Operational Technology (OT) Security; National Institute of Standards and Technology: Gaithersburg, MD, USA, 2023.
26. NIST. SP 800-207: Zero Trust Architecture; National Institute of Standards and Technology: Gaithersburg, MD, USA, 2020.
27. OWASP. Software Assurance Maturity Model (SAMM) v2.1; OWASP Foundation, 2023.
28. MITRE. ATT&CK for ICS Matrix; MITRE Corporation, 2023.
29. Efron, B.; Tibshirani, R.J. An Introduction to the Bootstrap; Chapman and Hall/CRC: New York, NY, USA, 1993.
30. Wilcoxon, F. Individual comparisons by ranking methods. *Biometrics Bulletin* 1945, 1, 80-83.



31. Cliff, N. Dominance Statistics: Ordinal Analyses to Answer Ordinal Questions; Psychological Bulletin 1993, 114, 494-509.
32. Forsgren, N.; Humble, J.; Kim, G. Accelerate: The Science of Lean Software and DevOps; IT Revolution Press: Portland, OR, USA, 2018.

Real-Time Classification of Hand Movements from Surface Electromyography Signals Using Feature Engineering and an Ensemble of Machine-Learning Classifiers for Rehabilitation

Turimov Dilmurod

Department of Computer Engineering, Gachon University, Sujeong-gu, Seongnam-si 461-701, Gyeonggi-do, Republic of Korea,

Department of Control Theory and Information Security, Samarkand State University named after Sharof Rashidov, Samarkand 140104, Uzbekistan

email: dilmurod@gachon.ac.kr

Kilichev Dusmurod

Department of Information Security, National University of Uzbekistan, Tashkent 100174, Uzbekistan,

email: dusmurod@nuu.uz

Saydirasulov Norkobil

Department of Computer Engineering, Gachon University, Sujeong-gu, Seongnam-si 461-701, Gyeonggi-do, Republic of Korea,

Department of Technologies of Mobile Communication, Tashkent University of Information Technologies Named After Muhammad Al-Khwarizmi, Tashkent 100200, Uzbekistan

email: saydirasulov@gachon.ac.kr

Mukhamadiyev Abdinabi

Department of Computer Engineering, Gachon University, Sujeong-gu, Seongnam-si 461-701, Gyeonggi-do, Republic of Korea,

email: mukhamadiyev@gachon.ac.kr

Abstract - Myoelectric control of prostheses and biosignal-driven rehabilitation require hand-movement recognition that is simultaneously accurate, computationally light, and fast enough for real-time operation on embedded hardware. This paper presents an end-to-end methodology for classifying object-interaction hand movements from surface electromyography (sEMG) signals. Four contributions are combined into a single deployable pipeline: (i) a real-time active-potential onset/offset detection algorithm based on rectified-signal variance thresholding (threshold = $\mu + h\sigma$); (ii) an optimal, low-dimensional feature vector obtained through feature engineering in the time, frequency and time-frequency domains; (iii) a real-time algorithm for automatically building the training set during signal acquisition; and (iv) an ensemble classifier that fuses k-Nearest Neighbours, Support Vector Machine and Random Forest decisions with a confidence gate and temporal confirmation. Two datasets were recorded - a single-channel BITalino dataset (20 subjects, 4 movements) and a five-channel BTS FreeEMG dataset (42 subjects, 7 movements). The proposed onset detector achieved the lowest mean error (15.45 samples) among six competing methods. A six-feature time-domain vector (SSI, ACC, IEMG, WL, RMS, MAV) reached up to

99% accuracy with a 223 ms response time, and the overall real-time accuracy for object-interaction movements was 94.6%. When embedded in a low-cost hardware-software complex and deployed for the rehabilitation of spinal-cord-injury (C5-C8) patients, the method increased the number of elicited hand movements by a factor of 2-3 and shortened the rehabilitation course by roughly threefold.

Keywords: surface electromyography; hand-movement classification; feature engineering; machine learning; ensemble classifier; real-time; myoelectric prosthesis; rehabilitation; spinal cord injury.

Introduction

The number of people whose motor function is fully or partially impaired continues to grow worldwide, driven by congenital disability, vascular and diabetic disease, and traumatic injury. According to the World Health Organization, tens of millions of people currently require prosthetic assistance, yet commercial myoelectric prostheses remain prohibitively expensive (on the order of tens of thousands of euros) while typically reproducing only two or three grasp patterns. This mismatch between clinical need and affordable, multifunctional technology motivates research into low-cost, sensor-efficient systems that recover as much hand function as possible.

Human movement originates in the motor cortex; descending impulses travel through the spinal cord to skeletal muscles, whose contraction generates the electromyography (EMG) signal. Decoding hand movements from this signal underpins both myoelectric prosthesis control and biosignal-based physiotherapy, and additionally supports the detection of neuromuscular pathologies. Surface EMG (sEMG) is attractive because it is non-invasive, but the raw signal is stochastic, low-amplitude (0.1-5 mV), band-limited (20-500 Hz), and highly sensitive to electrode placement, fatigue and motion artefacts. Consequently, raw sEMG cannot be classified directly with acceptable accuracy or latency; informative **feature engineering** and an appropriately chosen classifier are essential.

For any device that must respond to the user in real time, three requirements dominate: the **response time** (the sum of data-collection and data-analysis time), the **classification accuracy**, and the **computational cost** of the features and classifier. These requirements are frequently in tension: richer feature sets and larger sensor arrays can raise accuracy but inflate latency and hardware demands. The central question addressed in this work is therefore how to reach high accuracy for a useful repertoire of hand movements while using as few sensors and as light a feature set as possible.

This paper makes the following contributions, each validated experimentally and integrated into a single deployable system:

- 1) A real-time algorithm that determines the active-potential threshold of the EMG signal from the variance of the rectified signal, locating movement onset and offset more accurately than five established alternatives.
- 2) An optimal, low-dimensional feature vector derived through systematic feature engineering, which maximises classification accuracy while remaining cheap to compute on embedded hardware.
- 3) A real-time algorithm that automatically assembles a labelled training set during signal acquisition, removing the manual effort usually required to build sEMG corpora.
- 4) A real-time ensemble classifier that fuses k-NN, SVM and Random Forest decisions, embedded in a low-cost hardware-software complex and validated clinically on spinal-cord-injury patients.

The remainder of the paper is organised as follows. Section 2 reviews related work. Section 3 describes the two datasets, the signal-processing chain, the active-potential detector, the feature-engineering procedure and the ensemble classifier. Section 4 reports the experimental results, including a comparison with the state of the art. Section 5 presents the clinical deployment, Section 6 discusses the findings, and Section 7 concludes.

Related Work

EMG-based movement recognition has been studied extensively. Early myoelectric systems relied on artificial neural networks (ANN, BPNN) and their probabilistic variants (LLGMN, RLLGMN), as well as hidden Markov models and Bayesian networks. Reported accuracies range widely - from around 70% for cursor-control tasks using back-propagation networks to about 94-95% for specific architectures - but neural approaches generally require multi-layer structures, long training, careful stochastic-parameter tuning, and are sensitive to small mislabelled movements, which limits their suitability for latency-critical embedded control.

A parallel body of work has concentrated on **feature engineering**. Time-domain features such as mean absolute value (MAV), waveform length (WL), zero crossings (ZC), slope-sign changes (SSC), root-mean-square (RMS) and Willison amplitude (WAMP) are computationally cheap and well suited to real-time use, whereas frequency-domain and time-frequency features (power-spectral-density descriptors, autoregressive coefficients, wavelet transforms) capture complementary information at higher computational cost. Comparative studies have repeatedly identified WL, MAV and RMS as among the most discriminative and noise-robust individual descriptors, and the Hudgins set (MAV, WL, ZC, SSC) as an effective compact combination. Feature-selection and reduction techniques - the Davies-Bouldin index, the ReliefF algorithm and principal component analysis - are commonly used to prune redundant descriptors.

Compared with neural networks, classical machine-learning (ML) classifiers - k-Nearest Neighbours (k-NN), Support Vector Machines (SVM) and Random Forests (RF) - offer faster training, few hyper-parameters, and straightforward integration with pre-computed features, which is why they are increasingly preferred in myoelectric control. However, most publicly available sEMG datasets either lack sufficient movement repetitions, or are recorded in a single session, both of which undermine the reliability of the trained model. The present work addresses these gaps by combining a robust onset detector, an engineered compact feature vector, an automatically generated multi-session training set, and an ensemble classifier, and by validating the whole pipeline both offline and in a real embedded deployment.

Materials and Methods

The complete recognition pipeline comprises seven stages - signal acquisition, pre-processing and band-pass filtering, active-potential detection, feature extraction, feature selection, ensemble classification, and command actuation (Figure 1). Each stage is designed so that the cumulative data-collection and data-analysis time remains compatible with real-time control.

Real-time hand-movement recognition pipeline



Figure 1. Real-time hand-movement recognition pipeline. Shaded blocks denote the novel components proposed in this work.

Data acquisition and datasets

Two datasets were recorded from healthy volunteers who took part voluntarily and in a positive psychological state. Before each session, participants completed a questionnaire covering sex, age, weight, health history, smoking and sporting activity, and the forearm and elbow diameters were measured. Signals were recorded from the forearm muscles responsible for grasping using non-invasive Ag-Cl surface electrodes placed over the muscle innervation zones. Recording was carried out gradually to avoid stress and fatigue and to obtain clean, low-artefact signals. The two corpora are summarised in Table 1.

Property	Dataset I (single-channel)	Dataset II (multi-channel)
Device	BITalino (Plux, Portugal)	BTS FreeEMG
Channels	1 (extensor carpi radialis)	5 (innervation zones)
Subjects	20 healthy	42 healthy (5 F, 37 M), 19-37 y
Movements	4 (3 active + rest)	7 (6 active + rest)
Repetitions	20 per movement	up to 30 per movement
Sessions	2 (over 1 week)	2 (over 2 weeks)
Sampling rate	1 kHz	1 kHz
Band-pass filter	band-pass, 20-500 Hz	band-pass, 20-500 Hz
Window length	200 ms (sequential)	256 ms
Samples	1 600 (20×4×1×20)	multi-session, multi-subject

Table 1. Characteristics of the two recorded surface-EMG datasets.

The four movements of Dataset I - making a fist (grip), pinch grip with the fingers, opening the hand, and a rest (motionless) state - represent the primary object-interaction functions of the palm. Dataset II extends this to six active movements (finger grip, fist, hand opening, and flexion

of the index, middle and ring fingers) plus the rest state. Signals were split 85% for training and 15% for testing.

Pre-processing and active-potential detection

Raw sEMG was recorded in differential mode and band-pass filtered (low cut-off 20 Hz, high cut-off 500 Hz) with a sixth-order Butterworth filter to remove baseline drift, motion artefacts and out-of-band noise. Because features must be computed from the muscle-active portion of the signal, robust detection of activation onset and offset is critical for both accuracy and speed. The proposed detector operates on the rectified signal and compares the windowed variance against a data-adaptive threshold

$$threshold = \mu + h \cdot \sigma, \quad \sigma = \sqrt{(1/N) \sum (x_i - \bar{x})^2} \quad (1)$$

where μ and σ are the mean and standard deviation of the EMG values, h scales the threshold level, and N is the number of samples. Starting from the onset point, the loop advances until the windowed variance of the rectified signal exceeds the product of the threshold and the noise variance. In the resulting indicator signal, a 0→1 transition marks activation onset and a 1→0 transition marks offset. The detector was compared with the single-threshold, double-threshold, Komi, sign-change and Teager-Kaiser energy operator (TKEO) methods (Section 4.1).

Feature engineering and the optimal feature vector

Features were extracted from three domains: the time domain (TD), the frequency domain (FD) and the time-frequency domain (TFD). Each descriptor was first evaluated **individually** and then in **combinations**, and the most discriminative descriptors were retained. Candidate selection was guided by three criteria - class separability, robustness to noise, and low computational complexity - and supported by the Davies-Bouldin index, the ReliefF weighting algorithm and principal component analysis for dimensionality reduction.

Individual descriptors reached only 40-60% accuracy on average, with WL the strongest single feature (87% with k-NN, $k = 3$) and ZC the weakest (33%). Accuracy improved sharply once several energy-related time-domain descriptors were combined. The retained optimal vector consists of six descriptors defined in Table 2: simple square integral (SSI), average amplitude change (ACC), integrated EMG (IEMG), waveform length (WL), root-mean-square (RMS) and mean absolute value (MAV).

Feature	Definition	Physiological meaning
SSI	$\sum x_i^2$	Signal energy index
ACC	$(1/N) \sum x_{i+1} - x_i $	Mean waveform-length change
IEMG	$\sum x_i $	Summed activation over the window
WL	$\sum x_i - x_{i-1} $	Cumulative waveform length
RMS	$\sqrt{(1/N) \sum x_i^2}$	Amplitude / contraction level
MAV	$(1/N) \sum x_i $	Rectified mean; contraction degree

Table 2. The six-descriptor optimal time-domain feature vector (x_i = sample; N = samples per window).

Machine-learning classifiers and the ensemble strategy

Three classical supervised classifiers were used. **Random Forest (RF)** aggregates the votes of many decision trees, each trained on a bootstrap sample using a random subset of m features per node; it is efficient on large data, resists over-fitting and provides feature-importance estimates. **Support Vector Machine (SVM)** constructs the maximum-margin separating hyperplane, using kernels (radial-basis-function or polynomial) for non-linearly separable data and either a one-against-one or one-against-all strategy for the multi-class problem. **k-Nearest Neighbours (k-NN)** is a non-parametric rule that assigns the majority class among the k closest training samples under the Euclidean distance.

For deployment the three classifiers were combined into a real-time ensemble (Figure 1, final block). Each incoming 25 ms sliding-window feature vector is normalised ($z = (x - u)/s$) and passed to all three models; the movement command is issued only when the mean of their class probabilities exceeds 0.85 and the same class is confirmed over four consecutive windows ($h = 4$). This confidence gate and temporal confirmation suppress spurious activations, which is essential when the output drives a physical actuator. The training set required by these models is generated automatically during acquisition: once the device connection is established and the active region is detected, the onset-to-offset segment is written to the corpus with a fixed array size, ensuring uniform training samples.

Results

Active-potential detection performance

Table 3 and Figure 2 compare the six onset-detection algorithms in terms of mean error and mean deviation, together with the number of signal samples each method processes. The proposed variance-threshold detector achieved the lowest mean error (15.45) and the lowest mean deviation (18.9), clearly outperforming the single-threshold method (31.0 / 32.2) and the classical Komi algorithm (29.5 / 32.8). Accurate onset localisation matters because features extracted from the correct activation instant improve both the speed and the accuracy of the downstream classifier.

Algorithm	Mean error	Mean deviation	Samples
Proposed	15.45	18.9	352
TKEO	20.2	27.5	402
Double threshold	25.6	22.6	156
Sign changes	24.8	25.8	718
Komi	29.5	32.8	956
Single threshold	31.0	32.2	98

Table 3. Comparative accuracy of active-potential (onset) detection algorithms (lower is better).

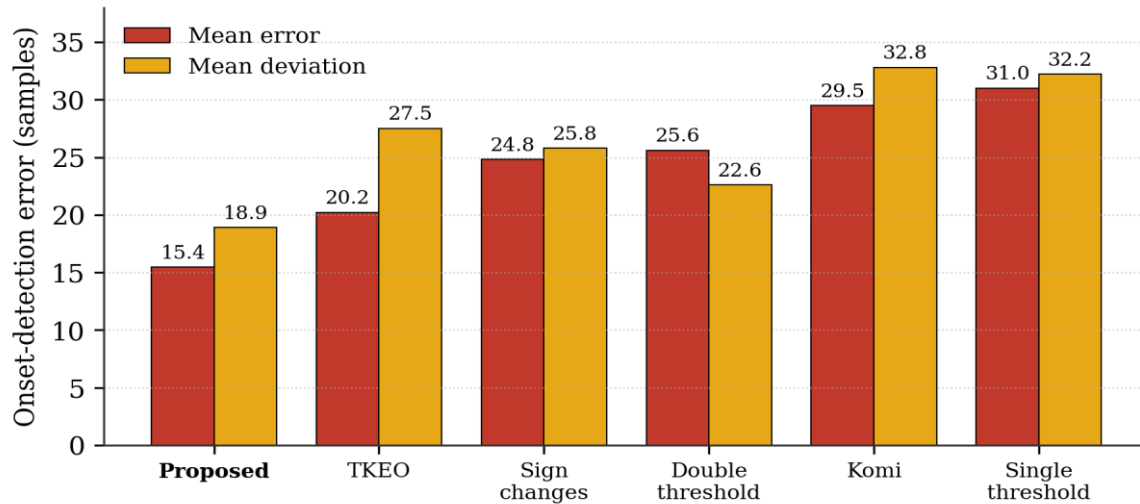


Figure 2. Mean error and mean deviation of the six onset-detection algorithms. The proposed method yields the smallest values.

Comparison of feature domains

Table 4 and Figure 3 summarise the mean classification accuracy and response time obtained with each feature domain on the four-movement task. Time-domain features gave the highest accuracy (95%) at the shortest response time (223 ms), of which 200 ms is the sequential-segmentation window and 23 ms is spent on pre-processing, feature extraction and classification. Frequency-domain and time-frequency features achieved 91% and 93% respectively but at two-to-four times the latency, because their computation is markedly more expensive - a decisive disadvantage for real-time control.

Feature domain	Mean accuracy (%)	Response time (ms)	Rank
Time domain (TD)	95	223	1
Time-frequency (TFD)	93	708-856	2
Frequency domain (FD)	91	443-556	3

Table 4. Mean accuracy and response time by feature domain (four-movement task).

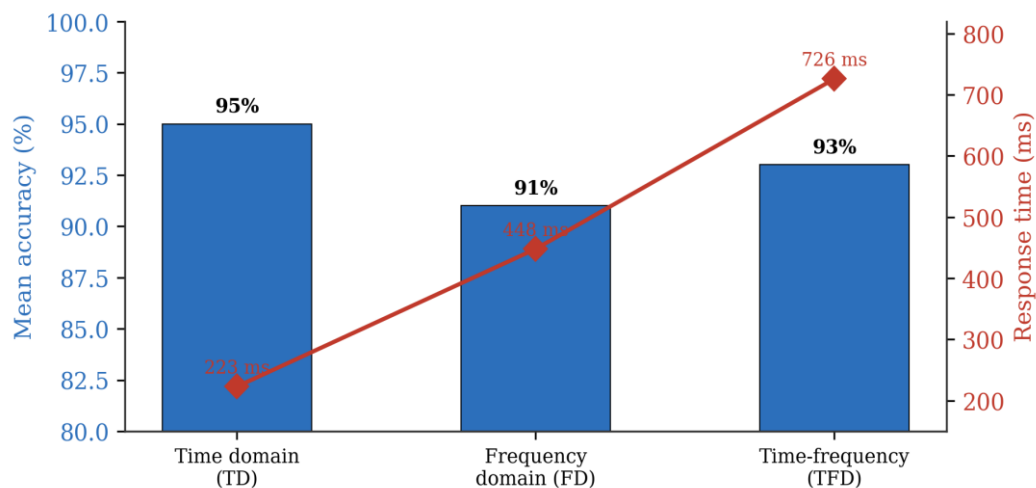


Figure 3. Trade-off between accuracy (bars, left axis) and response time (line, right axis) across the three feature domains.

Optimal feature-vector combination

Table 5 and Figure 4 trace how accuracy evolves as descriptors are added to the vector. Combining two features already lifts accuracy above the individual level, and results improve steeply from five features onward. The best result is obtained with the six-feature combination SSI + ACC + IEMG + WL + RMS + MAV, reaching 99% with k-NN, 99% with RF and 95% with SVM. Adding further descriptors (ZC, WAMP, log-detector) does not improve - and occasionally degrades - accuracy while increasing computational cost, confirming the six-feature vector as optimal.

Feature combination	k-NN (%)	RF (%)	SVM (%)
SSI + ACC	71	73	56
SSI + ACC + IEMG	86	87	70
SSI + ACC + IEMG + WL	94	97	89
SSI + WL + RMS	97	96	91
SSI + ACC + IEMG + WL + RMS	95	97	92
SSI + ACC + IEMG + WL + RMS + MAV	99	99	95
+ ZC	97	98	87
+ ZC + WAMP	98	95	89
+ ZC + WAMP + logDetect	98	97	90

Table 5. Classification accuracy as features are progressively combined; the highlighted six-feature vector is optimal.

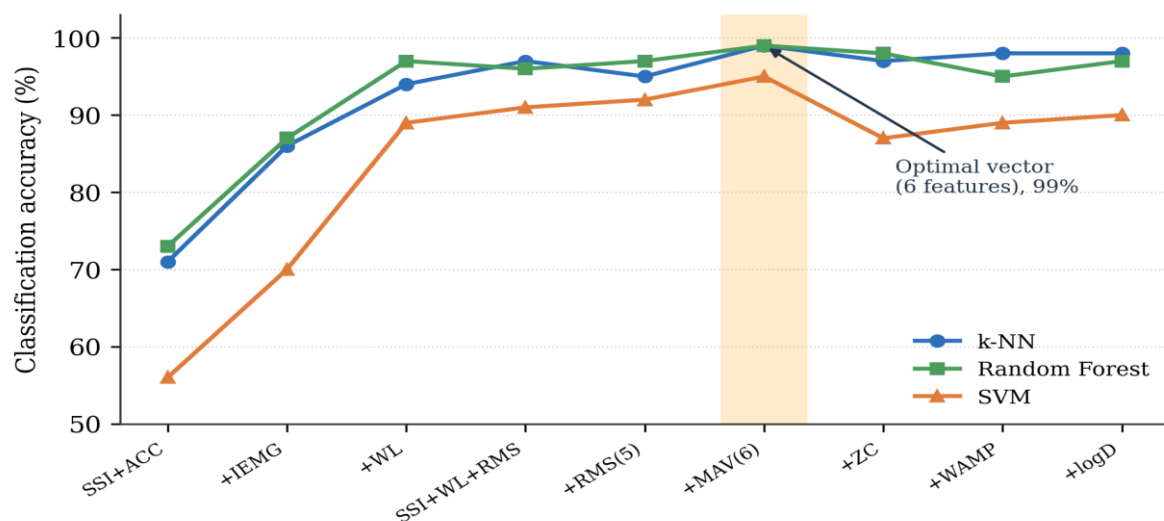


Figure 4. Accuracy progression of the three classifiers as descriptors are added. The six-feature vector attains 99%.

Effect of movement count and sensor count

Because Dataset I uses a single channel, the relationship between the number of target movements and accuracy was quantified (Figure 5a): a single-channel device can reliably classify at most three-to-four movements before accuracy falls below a usable level. On the multi-channel task (Figure 5b), accuracy peaks around five sensors and then plateaus or even decreases, because additional channels enlarge the real-time data stream and the feature-computation load. Random Forest slightly outperformed k-NN on the six-movement multi-channel task (94% versus 92% mean accuracy), and time-domain features again gave the highest per-movement accuracy. Across the three domains the overall figures were TD 95%, FD 91% and TFD 93%.

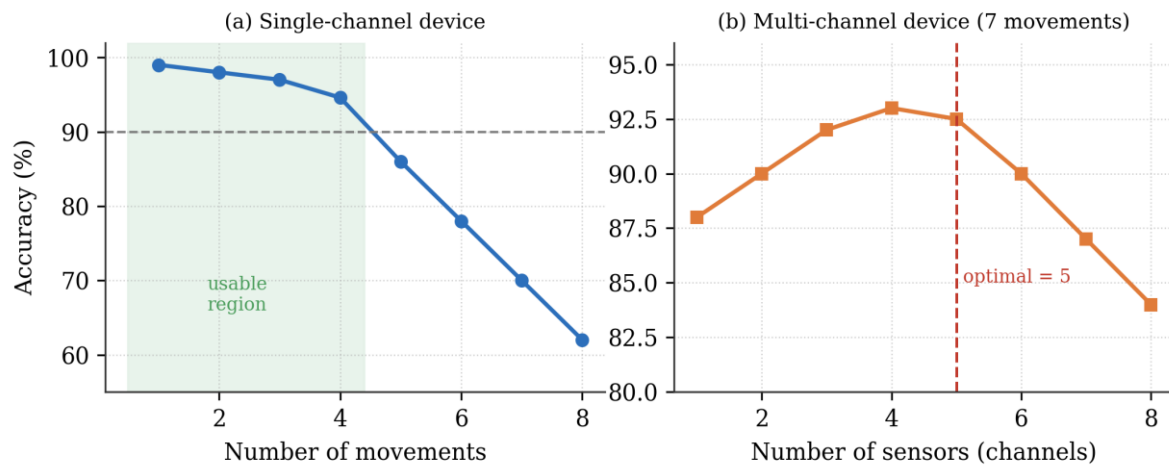


Figure 5. (a) Accuracy versus number of movements for a single-channel device; (b) accuracy versus number of sensors for the multi-channel device (seven movements).

Comparison with the state of the art

Table 6 places the two configurations of this work (marked * and) alongside twelve previously published systems. The comparison uses the number of movements, the number of sensors, the summed data-collection and data-analysis time (DCT + DAT), and the offline and real-time accuracies. Both configurations were tested on embedded systems** rather than on a PC. Using only one sensor for four movements, the proposed system reached 98% offline and 94.6% real-time accuracy at 223 ms; using five sensors for seven movements it reached 99% offline and 92.5% real-time at 232 ms. The distinguishing feature of this work is the deliberate minimisation of DCT + DAT and sensor count while keeping real-time accuracy above most reported systems.

Study	Movements	Sensors	DCT+DAT (ms)	Offline (%)	Real-time (%)	Platform
[1]	5	8	1029.4	90.7	85.3	PC
[2]	6	8	137.9	99.0	89.6	PC
[6]	10	4	250.2	94.0	85.0	Embedded
[7]	7	8	300.0	97.9	83.0	PC
[9]	9	8	254.5	99.0	90.2	Embedded
[12]	8	32	205.0	98.2	92.3	Embedded

Study	Movements	Sensors	DCT+DAT (ms)	Offline (%)	Real-time (%)	Platform
This work *	4	1	223	98.0	94.6	Embedded
This work **	7	5	232	99.0	92.5	Embedded

Table 6. Comparison with representative published sEMG hand-recognition systems (selected rows; DCT + DAT = data-collection + data-analysis time).

Clinical Application

The hardware-software complex was designed around low-cost, widely available components (a BITalino acquisition board with an ATMEGA328P-AU micro-controller and a Bluetooth 2.0 link), a small number of sensors, and long battery life. It comprises five modules: signal acquisition, pre-processing, feature extraction, movement classification, and on-device testing. Beyond prosthesis control, the same system supports the rehabilitation of patients whose motor pathway is disrupted - most notably spinal-cord-injury (SCI) patients.

SCI arises predominantly from road accidents ($\approx 39\%$) and falls ($\approx 32\%$); injuries in the C5-C8 segments can disable the arm down to the shoulder, and the treatment course ranges from one month to several years at very high cost. In conventional physiotherapy a therapist forces muscle contraction with an electro-stimulator, typically eliciting only two movements. In the proposed method (Figure 6), the therapist's own hand movements are classified from EMG and the resulting commands drive stimulation of the patient's muscles, so the number of elicited movements scales with the number of acquisition channels.

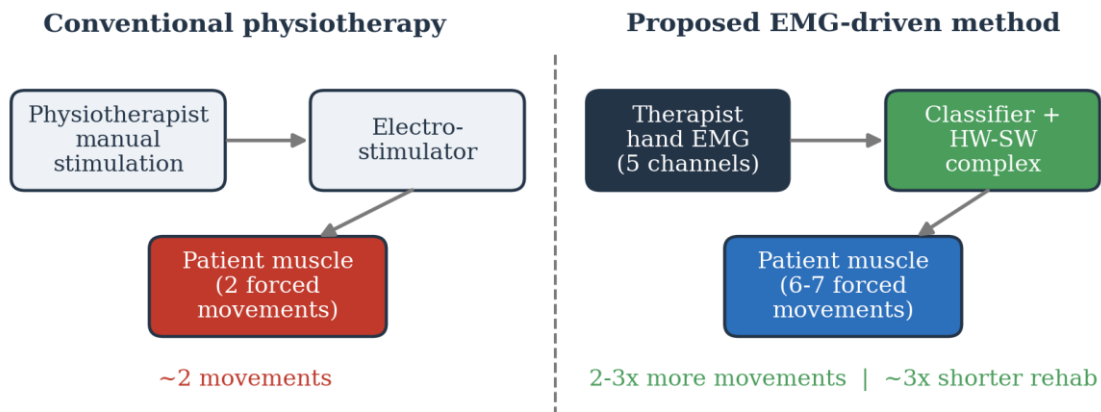


Figure 6. Conventional physiotherapy (left) versus the proposed EMG-driven rehabilitation scheme (right).

The method was deployed at the Kashkadarya Regional No. 1 Multidisciplinary Hospital and at the Kashkadarya Regional Rehabilitation and Physiotherapy Centre, and tested on SCI patients with C5-C8 injuries. In these trials the number of movements that could be elicited in the affected hand increased by a factor of two-to-three, giving the physiotherapy protocol greater multifunctionality, and the rehabilitation course was shortened by approximately threefold.

Discussion

The results support three main observations. First, accurate onset detection is a prerequisite for both speed and accuracy: by extracting features from the correctly localised activation segment, the proposed variance-threshold detector reduces onset error by roughly half relative to the single-threshold and Komi baselines. Second, for real-time embedded control, a compact engineered time-domain vector is preferable to richer frequency or time-frequency representations; the six-feature vector matches or exceeds their accuracy at a fraction of the latency. Third, sensor count must be matched to the movement repertoire - more channels are not automatically better, since they increase computation and can even reduce accuracy, with roughly five sensors sufficient for multifunctional operation and one sensor adequate for three-to-four movements.

Fusing k-NN, SVM and RF with a confidence gate and four-window temporal confirmation trades a small, bounded latency for markedly more stable output, which is critical when the classifier drives a physical actuator or a patient's muscle. The automatic, multi-session training-set generator further improves reliability by capturing the physiological variability and fatigue that single-session corpora miss. The principal limitations are that both datasets were recorded from healthy subjects and that clinical validation, while promising, involved a limited patient cohort; larger multi-centre trials and evaluation on amputees are natural next steps, together with closed-loop sensory feedback for the myoelectric-prosthesis use case.

Conclusion

This paper presented a complete, deployable methodology for real-time hand-movement classification from surface EMG. Its four integrated contributions - a variance-threshold active-potential detector (lowest mean error, 15.45), an optimal six-feature time-domain vector (up to 99% accuracy), a real-time automatic training-set generator, and a confidence-gated k-NN/SVM/RF ensemble - together achieved 94.6% real-time accuracy for object-interaction movements at a 223 ms response time on embedded hardware. Embedded in a low-cost hardware-software complex and deployed for SCI rehabilitation, the system increased the number of elicited hand movements 2-3-fold and shortened the rehabilitation course roughly threefold. By jointly optimising accuracy, latency and sensor count, the approach offers an affordable route to multifunctional myoelectric control and biosignal-driven physiotherapy. Future work will extend validation to amputees and larger patient cohorts and integrate sensory feedback.

Reference

1. Chen, Z., Min, H., Wang, D., Xia, Z., Sun, F., & Fang, B. (2023). A review of myoelectric control for prosthetic hand manipulation. *Biomimetics*, 8(3), 328.
2. Jiang, X., Liu, X., Fan, J., Ye, X., Dai, C., Clancy, E. A., Farina, D., & Chen, W. (2022). Optimization of HD-sEMG-based cross-day hand gesture classification by optimal feature extraction and data augmentation. *IEEE Transactions on Human-Machine Systems*, 52(6), 1281-1291.
3. Miah, A. S. M., Shin, J., & Hasan, M. A. M. (2024). Effective features extraction and selection for hand gesture recognition using sEMG signal. *Multimedia Tools and Applications*, 83, 1-25.
4. Aarotale, P. N., et al. (2024). Machine learning-based sEMG signal classification for hand gesture recognition. *arXiv preprint arXiv:2411.15655*.

5. Zhang, R., Zhang, X., He, D., Wang, R., & Guo, Y. (2022). sEMG signals characterization and identification of hand movements by machine learning considering sex differences. *Applied Sciences*, 12(6), 2962.
6. Junior, J. J. A. M., Pontim, C. E., Dias, T. S., & Campos, D. P. (2023). How do sEMG segmentation parameters influence pattern recognition process? An approach based on a wearable sEMG sensor. *Biomedical Signal Processing and Control*, 81, 104546.
7. Lv, X., et al. (2023). Gesture recognition based on sEMG using a multi-attention mechanism for remote control. *Neural Computing and Applications*, 35(19), 13839-13849.
8. Lee, K. H., Min, J. Y., & Byun, S. (2022). Electromyogram-based classification of hand and finger gestures using artificial neural networks. *Sensors*, 22(1), 225.
9. Trajdos, P., & Kurzynski, M. (2024). Application of context-dependent interpretation of biosignals recognition to control a bionic multifunctional hand prosthesis. *Biocybernetics and Biomedical Engineering*, 44(1), 161-182.
10. Trajdos, P. (2025). A noise-tolerant dual multi-classifier system with fuzzy model applied to the sEMG-based control of a bionic upper limb prosthesis. *Biomedical Signal Processing and Control*, 104, 107441.
11. Iqbal, H., Zheng, J., Chai, R., & Chandrasekaran, S. (2024). Electric powered wheelchair control using user-independent classification methods based on surface electromyography signals. *Medical & Biological Engineering & Computing*, 62(1), 167-182.
12. Bicer, Y., et al. (2024). User training with error augmentation for sEMG-based gesture classification. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 32.
13. Dai, Y., et al. (2022). MSEva: a musculoskeletal rehabilitation evaluation system based on EMG signals. *ACM Transactions on Sensor Networks*, 19(1), 1-23.
14. Mokri, C., Bamdad, M., & Abolghasemi, V. (2022). Muscle force estimation from lower limb EMG signals using novel optimised machine learning techniques. *Medical & Biological Engineering & Computing*, 60(3), 683-699.
15. Zhou, P. (2024). Surface EMG in China: a report on the 2023 surface EMG symposium. *Frontiers in Rehabilitation Sciences*, 5, 1353564.