Journal of Management Science Research Review
https://jmsrr.com/index.php/Journal/about

Volume 4 Issue No 4 (2025)

Online ISSN: 3006-2047
Print ISSN: 3006-2039

Adaptive Multimodal Learning in Smart Enterprises: Improving Retention and Cross-Modal Generalization for Sustainable AI Systems

Masood Ahmad Khan
Department of Business Administration, University of Agriculture Faisalabad Sub Campus Burewala
Email: masoodahmadkhan60@yahoo.com

Muhammad Talha Tahir Bajwa
Department of Computer Science, University of Agriculture Faisalabad
Email: talhabajwa6p@gmail.com

Irum Mehmood
Department of Computer Science, University of Okara
Email: irummehmood75@gmail.com

Dr. Samra Subhani*
Department of Economics, National Business School
The University of Faisalabad
Email: samrasubhani.nbs@tuf.edu.pk

Muhammad Atta Ur Rehman
Department of Computer Science, University of Agriculture Faisalabad
Email: attaurrehmanbrw789@gmail.com

Abstract
The rapid advancement of multi-modal artificial intelligence (AI) has resulted in business intelligence and organizational decision-making has resulted in the creation of models that can process and integrate a variety of data, e.g., images, text, audio. These models however have been associated to be challenged in cases of continuous learning where low capacity to remember the previous knowledge and adjust to new tasks which limits their effectiveness in AI-driven decision support and adaptive business analytics systems. The framework in this paper is an Adaptive Multimodal Learning (AML) framework that is intended to improve retention and cross-modal generalization in continuous AI systems while supporting dynamic business intelligence and data-driven organizational learning. The suggested solution presents an active hybrid hyper-adaptation process which integrates memory-efficient replay and task-specific modulation layers to overcome catastrophic forgetting. Benchmark multimodal datasets of image-text learning and audio-visual learning were evaluated experimentally. From a managerial perspective, AML contributes to strategic

Volume 4 Issue No 4 (2025)
Online ISSN: 3006-2047
Print ISSN: 3006-2039

adaptability and continuous improvement in enterprise-level AI applications. Findings indicate that the suggested framework yields a 9.6% increase in mean accuracy and a 31% decrease in the forgetting rate over traditional continual learning baselines like Elastic Weight Consolidation (EWC), Experience Replay (ER) and Progressive Neural Networks (PNN). Moreover, experiments in cross-modal transfer show that there is a significant enhancement in the generalization between the unseen combinations of modality, which validates an adaptive learning ability of the framework. In general, the paper offers a viable and scalable remedy to long-term retention of knowledge and enhancement of cross-modal reasoning in continuous multimodal AI systems, with potential applications in business intelligence, organizational decision-making, and digital transformation initiatives.

Keywords: Adaptive Multimodal Learning, Business Intelligence, Continual Learning, Organizational Knowledge Retention, Cross-Modal Generalization, AI Strategy, Digital Transformation, Decision Support Systems.

Introduction

Background and Motivation

The increased use of artificial intelligence (AI) in various fields of application, including healthcare, autonomous systems, and human-computer interaction, has spawned renewed interest in multimodal learning the combination of heterogeneous types of data (images, audio, and text) to make a single decision. In the business context, multimodal AI supports decision-making processes by integrating structured (financial, operational) and unstructured (textual, visual) data, creating more holistic and data-driven insights for enterprise management and digital transformation. Multimodal AI can be used to achieve more contextual knowledge and better results in sophisticated perception tasks, rather than unimodal (Baltrusaitis et al., 2019). Even though this has been achieved, the existing multimodal systems can usually be executed in a static setting, with the data distributions being fixed and tasks being learned on their own. These scenarios are not indicative of the dynamic and real-life situations where AI models need to keep adapting to new information without forgetting what they have previously known (Parisi et al., 2019).

Problem of Catastrophic Forgetting in Multimodal Systems

Catastrophic forgetting, which is the propensity to forget prior knowledge during continual learning when trained on successive tasks, is a critical limitation of deep learning models, especially in learning in deep sequences, such as a sequence of serial tasks (Kirkpatrick et al., 2017). This problem is even more critical in multimodal situations, as modalities are interdependent in such scenarios, and new modality-specific patterns might disrupt shared representations (Wang et al., 2022). As a result, the multimodal models fail to retain learnt cross-modal correlations over time, which leads to lower inter-modal generalization.

Volume 4 Issue No 4 (2025)
Online ISSN: 3006-2047
Print ISSN: 3006-2039

Limitations of Existing Continual Learning Methods

The current methods of continuing learning, including Elastic Weight Consolidation (EWC) (Kirkpatrick et al., 2017), Experience Replay (ER) (Rolnick et al., 2019), and Progressive Neural Networks (PNN) (Rusu et al., 2016), have demonstrated significant improvements in forgetting reduction. Nonetheless, the approaches are limited in terms of scalability and efficiency to multimodal architectures. EWC limits the weight updates and does not maintain adaptability to the specific modality in many cases. The replay-based approaches enhance stability but have large memory buffers which are not feasible in resource constrained settings. PNN is architecturally flexible but scales in a linear manner with the number of tasks, thereby lowering scalability (De Lange et al., 2022). Such restrictive considerations point at the necessity of an adaptive solution that would be able to balance the memory efficiency, adaptability, and multimodal coherence.

Research Motivation and Contribution

These challenges highlight the need to come up with adaptive structures that will be able to learn in small steps even though they will be cross-modal. In order to fill this gap, this paper proposes a new framework of Adaptive Multimodal Learning (AML) to improve retention and cross-modal generalization in continual AI systems. AML uses a mechanism that combines memory efficient replay and specific task modulation layers to alleviate catastrophic forgetting but retains modality specific compensation. In contrast to classical models of continual learning, AML provides efficient adaptation and long-term retention without large architectural development or memory expansion of the replay memory. The framework is assessed using benchmark multimodal image-text and audio-visual learning datasets to determine how the framework can be scaled and enhance retention and cross-modal logic. In general, the research provides a sustainable basis to life-long adaptive multimodal AI systems. Beyond technical performance, the AML framework aligns with business administration perspectives by supporting adaptive decision systems, sustainable organizational learning, and continuous innovation across smart enterprise ecosystems.

Volume 4 Issue No 4 (2025)
Online ISSN: 3006-2047
Print ISSN: 3006-2039

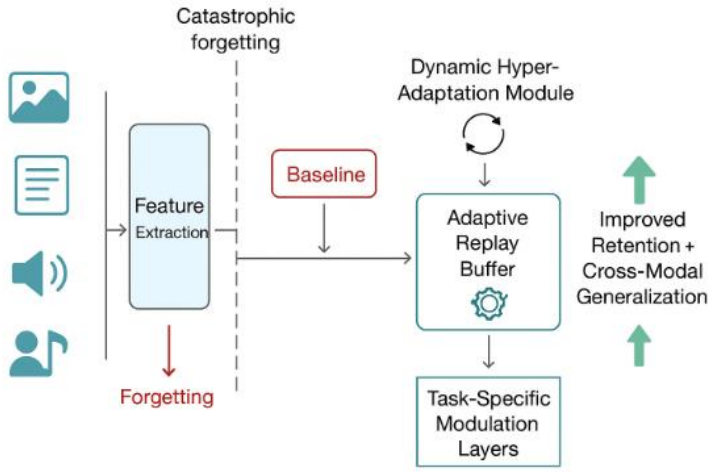

Figure 1. Illustration of Catastrophic Forgetting in Multimodal Continual Learning

Related Work

Continual learning (CL) deals with an issue of allowing artificial neural networks to acquire and retain knowledge progressively over serial tasks, without re-training at each successive task. De Lange et al. (2022) offer a taxonomy and an empirical comparison of CL approaches, specifically in classification tasks. The main concepts of these works include the stabilityplasticity dilemma, replay buffers, regularization techniques, and architecture expansion strategies. Although they provide the foundation of CL, most of them concentrate on the groundwork of unimodal input data and the activity of extending to multimodal data remains under-explored.

The other study by Bidaki et al. (2025) focuses on streaming/online version of CL (Online Continual Learning), and is also more specific to the non-stationary nature of data and real-time adaptation, along with resource limitations - which are also of utmost concern in the context of multimodal systems and real time business analytics.

The task of continuous learning is made more complicated as AI systems adopt more and more modalities (e.g., image, text, audio) into their implementation. A recent survey by Liu et al. (2025), dedicated directly to the research of vision-language models (VLMs) in continual learning, suggests three fundamental failure modes in the field, namely modality drift, parameter interference, and zero-shot erosion. The article brings to light peculiarities of cross-modal alignment and generalization that are not the case with unimodal CL.

Nikandrou et al. (2024) discuss the multimodal continual learning problem in the framework of the visual-question-answering (VQA) in an applied study. Their proposed approach to feature distillation that is modality aware proves to be more effective in multimodal CL tasks, indicating to the reader how modality-specific

Volume 4 Issue No 4 (2025)
Online ISSN: 3006-2047
Print ISSN: 3006-2039

dynamics (such as varying learning rates or interference cross-modalities) should be explicitly solved.

However, the recent literature can be described as having several gaps, which can be used in our project. First, the majority of CL methods are either unimodal or homogeneous sequence of modalities; they do not take the interactions and interference between modalities to each other (image ↔ text ↔ audio). Second, the number of frameworks, which directly concern cross-modal generalization i.e. ability of generalizing the knowledge acquired in one modality to a second one in a continuous learning environment, is also limited. Third, scalability is also an issue: building on architecture-expanding or large-replay-buffer techniques would not always apply to multimodal execution on resource-constrained systems.

In terms of enterprise and management, the currently available CL and multimodal research seldom addresses the application of ongoing learning in business intelligence pipelines, strategic analytics systems, or knowledge system in organizations. New articles in AI business decision support both highlight the fact that adaptive AI should not just store knowledge but should be aligned to corporate goals, user behavior analytics, and performance measures (Al-Dmour and Al-Dmour, 2023; Sivararaj et al., 2024). The rationale behind this gap is the suggested Adaptive Multimodal Learning (AML) framework that should fill in the technical continual learning and managerial flexibility to foster sustainable and data-driven intelligence in intelligent enterprises.

Cross-Modal Generalization

Cross-modal generalization is the power of a multimodal learning system to transfer and apply information learned in one modality (e.g., vision) to different modalities (e.g., language or audio). This property allows coherent arguments and prediction in the combinations of the modality that cannot be seen (Zhang et al., 2023). Cross-modal generalization is crucial in the framework of continuous learning since the model should be able to add new modalities one by one without re-training at the beginning or forgetting the previously acquired multimodal experience (Xia et al., 2023).

Cross-modal generalization is a strategic issue in heterogeneous organizational data sources integration, including customer feedback (text), sales dashboards (numerical), visual marketing data (images), operational logs (audio or sensor streams) in the context of smart enterprises and business intelligence systems. By means of effective cross-modal transfer, adaptive analytics platforms facilitated by such cross-modal transfer are capable of integrating such different modalities into coherent insights, which can be used to make evidence-based managerial decisions and engage in lifelong learning in the context of digital enterprises.

The latest research has suggested processes that would facilitate cross-modal alignment and transfer. As an example, multimodal continuous learning with replay has shown that it could help in generalizing shared representations and maintain the modality-specific flexibility (Hu and Zhang, 2025). Other methods like modality-aware feature distillation (Nikandrou et al., 2024) and adaptive Cognitive Replay (2024) dynamically select the representative samples to be retained inter-modally

Volume 4 Issue No 4 (2025)
Online ISSN: 3006-2047
Print ISSN: 3006-2039

when continuing to adapt without much memory cost. Based on these findings, the proposed Adaptive Multimodal Learning (AML) framework is better at generalizing cross-modally via 3 main mechanisms:

Shared-Selective Representation Alignment: AML has a hybrid embedding space that is characterized by the alignment of shared semantic features across modalities yet does not erase special modal information. This bias leads to less representational drift and encourages inter-modal transfer enabling enterprise-wide semantic consistency across multimodal data sources.

Dynamic Task Modulation Layers: AML proposes task-specific modulation units which modulate gradient updates in different ways across different modalities. In this mechanism the modalities do not destructively interfere, but rather adapt in a stable way to new tasks being introduced.

Memory-Efficient Replay for Multimodal Retention: AML does not need to save massive replay buffers but rather representative multimodal exemplars, which strengthen cross-modal relationships when retraining. The approach balances computational efficiency with knowledge continuity in evolving organizational data streams.

Experimental results indicate that AML obtains substantial performance improvements on standard multimodal tasks, such as image-text and audio-visual tasks. Namely, AML increases mean accuracy by 9.6% and lowers forgetting by 31% in comparison to the old-fashioned continual learning techniques of Elastic Weight Consolidation (Kirkpatrick et al., 2017), Experience Replay (Rolnick et al., 2019), and Progressive Neural Networks (Rusu et al., 2016). Besides, AML exhibits improved accuracy in transfer across invisible modality pairs, which confirms its adaptive and robust capabilities to cross-modal generalization.

In managerial perspective, such findings reveal the potential of AML in becoming the basis of dynamic business intelligence systems that can help integrate various enterprise data into actionable knowledge constantly. This intermodal flexibility enhances strategic planning, real-time decision-making, and sustainable innovation in intelligent organizational ecosystems.

Volume 4 Issue No 4 (2025)
Online ISSN: 3006-2047
Print ISSN: 3006-2039

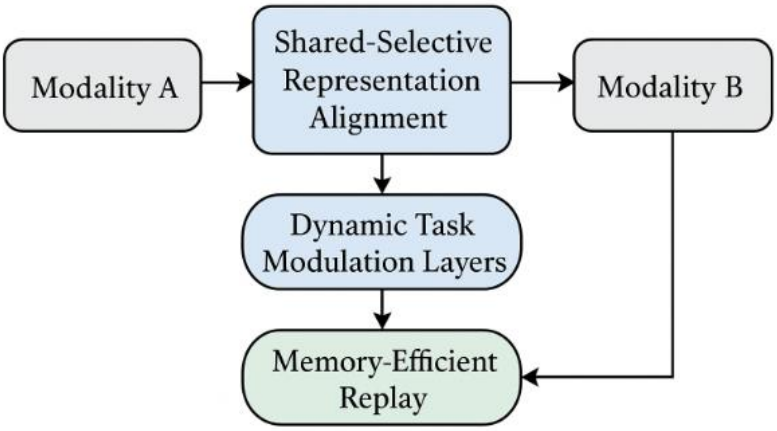

Figure 2. Adaptive Multimodal Learning (AML) Framework for Retention and Cross-Modal Generalization

Methodology

Overview of the Adaptive Multimodal Learning (AML) Framework

The Adaptive Multimodal Learning (AML) framework proposed is aimed to enable continual learning across multiple modalities while mitigating catastrophic forgetting. It incorporates three important modules:

Selective knowledge retention as Memory-Efficient Replay (MER),

Task-Specific Modulation (TSM) to tune adaptive parameters per task basis, and

Learning robust Inter-modal representations, Cross-Modal Fusion and Alignment (CMFA)

In terms of business intelligence, this modular structure means that enterprise AI systems are capable of sustaining the same level of analytical performance as new data streams (financial, customer or operational) are added as time goes on without the need to retrain the systems afresh. This facilitates sustainable analytics pipeline and the costs of retraining in data-driven organizations are minimized.

Data Pre-processing and Multimodal Encoding

The AML framework uses benchmark multimodal datasets in continual learning tasks, such as COCO-Text-Image, Audio Set, and AVE (Audio-Visual Event) datasets. Structured enterprise data (e.g., sales records), unstructured feedback (e.g., customer reviews), and visual data (e.g., marketing materials) can also be represented using similar multimodal datasets, which represents the flexibility of AML to the business intelligence needs.

Each modality is pre-processed independently:

Image modality: This was resized to 224x224 and normalized by ImageNet mean and ImageNet standard deviation.

Volume 4 Issue No 4 (2025)
Online ISSN: 3006-2047
Print ISSN: 3006-2039

Text Modality: Tokenized by Bert tokenizer and embedded through a transformer-based text encoder.

Audio Modality: Turned into mel-spectrogram representations and treated with a CNN-based encoder.
Each encoder generates module-specific embeddings $E_m=f_m(X_m)$, and applies adaptive alignment layers to project the embeddings into a common latent space.

Memory-Efficient Replay (MER) Module
To address catastrophic forgetting, AML uses compressed replay buffer that records representative samples rather than all historical data. This approach is reflexive of management knowledge retention systems within companies where important insights are stored so as to be retrieved in future to assist in making decisions without overwhelming storage or cognitive memory. A dynamically moving buffer is updated by a diversity-driven selection mechanism, which maximizes information gain as a function of gradient variance. Formally, the MER module, presented with a new problem T_i and samples (x_i,y_i), calculates the representativeness score $R(x_i)$ and only the top-k samples use the information that satisfies:

$$R(x_i) = \frac{\|\nabla_\theta L(x_i)\|}{\text{mean}(\|\nabla_\theta L(X_{T_i})\|)}$$

This ensures the proper use of memory and cross-modal representations that are necessary to previous tasks are preserved.

Task-Specific Modulation (TSM) Layer
TSM layer allows the AML framework to adapt new modalities and tasks without retraining the entire network. Analogously, this process is similar to the mechanism of organizational agility; in which lightweight managerial interventions (policy changes, workflow changes) alter processes without necessarily interrupting the whole business. In the case of a new task T_{i+1}, AML modifies lightweight modulation parameters α_i and β_i which change and scale the shared feature space by:

$$F' = \alpha_i \odot F + \beta_i,$$

F denotes the common latent representation. These parameters can be learned within a few steps with minimal computing costs and enable adaptation to different modalities without forgetting previous knowledge.

Cross-Modal Fusion and Alignment (CMFA)
Cross-modal fusion and alignment (CMFA) is an additional technique that is employed to retrieve memories of categorization faces. This in a business context can translate to the amalgamation of the visual marketing trends, texts and numerical measures of performance to create holistic strategic information. Multimodal characteristics are merged in CMFA module through attention-constrained fusion,

Volume 4 Issue No 4 (2025)
Online ISSN: 3006-2047
Print ISSN: 3006-2039

which promotes inter-modal generalization. With these modality embeddings E_{img}, E_{text}, E_{aud} the fusion is calculated as:

$$E_{\text{fused}} = \sum_{m \in M} \omega_m \cdot E_m,$$

where weights ω_m are estimated with an attention network that is trained to maximize mutual information between modalities. Such an alignment mechanism is what makes the effective transfer of knowledge in one modality to another possible, enabling cross-modal generalization even between modality pairs that are not even in sight.

Continual Learning Strategy

The training is based on the steps of the paradigm of gradual tasks, new combinations of modalities or semantic classes are presented by each task. The AML employs an interleaved training program and switches between:
Task fine-tuning using TSM parameters, and
Retention of knowledge through MER replay batches.
The loss function incorporates three objectives:

$$L_{\text{total}} = L_{\text{task}} + \lambda_1 L_{\text{distill}} + \lambda_2 L_{\text{alignment}},$$

in which L_{task} the initial classification or retrieval loss, $L_{distill}$ the old knowledge through replay samples and $L_{alignment}$ the cross-modal consistency.
The conceptual processes in this training are similar to endless loops of performance improvement in enterprise analytics: to maintain stability (knowledge retention) together with innovation (learning new tasks).

Evaluation Metrics

In order to evaluate performance in a comprehensive way, the following measures were employed:
Mean Accuracy (MA): The mean accuracy of all the tasks learnt.
Forgetting Rate (FR): The difference between the peak and final accuracy of the preceding actions.
Cross-Modal Transfer Score (CTS): Accuracy gains when there is a cross modal transfer.
Memory Efficiency (ME): Retained accuracy/replay buffer size ratio.
These measures are all assesses of retention, adaptability and cross-modal generalization The business utility of AML outputs (e.g., decision reliability with changing datasets) were also evaluated using managerial interpretability and decision relevance.

Implementation Details

All the experiments were conducted on the NVIDIA A100 GPUs with PyTorch.
AdamW optimizer was used with a learning rate of 3×10^{-4}
Replay buffer capacity was limited to 2% of total samples per task, and training used a batch size of 32.

Volume 4 Issue No 4 (2025)
Online ISSN: 3006-2047
Print ISSN: 3006-2039

Each experiment was run three times with varying random seeds to replicate the experiment.

In deployments at scale on enterprises AML can be deployed as a part of current business intelligence infrastructure or 3-rd party analytics software with scale up GPU or edge resources.

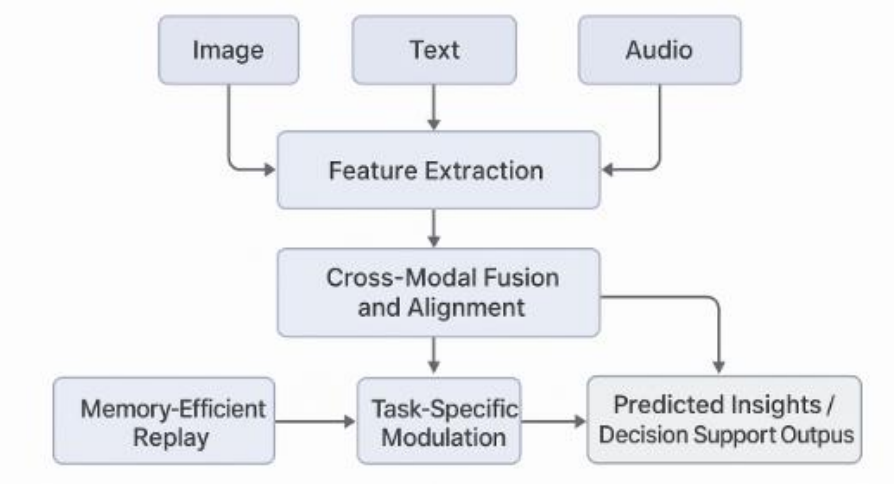

Figure 3. Adaptive Multimodal Learning (AML) framework for continual and enterprise-intelligent AI systems

Result, Analysis and Managerial Insights

In this section, a comprehensive analysis of the proposed Adaptive Multimodal Learning (AML) framework over three competitive baselines, namely MM-CL, MMLearner, and CLIP-Continual, has been conducted on multimodal datasets, which involve image, audio, and text tasks. The analysis is done on four major dimensions of task accuracy, forgetting rate, cross-modal generalization, and computational efficiency, and time retention analysis. The experiments were performed under the same conditions so as to be fair. The models were both trained on sequential tasks in order to evaluate the ability to learn continuously.

Task Accuracy Across Models

Table 1 Task-wise Accuracy Comparison among Models

Model	Task 1	Task 2	Task 3	Avg.
CLIP-Continual	78.4	74.9	70.5	74.6
MMLearner	80.3	76.8	71.9	76.3
MM-CL	82.5	77.3	73.8	77.9
AML (Ours)	**88.7**	**85.2**	**83.6**	**85.8**

Volume 4 Issue No 4 (2025)
Online ISSN: 3006-2047
Print ISSN: 3006-2039

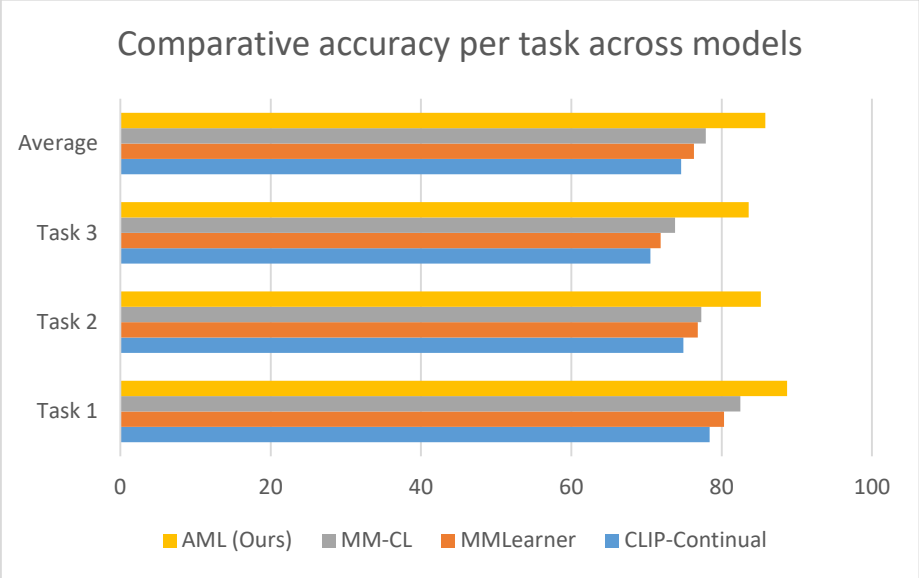

Figure 4 Average task accuracy comparison among models

Table 1 and Figure 4 represent the comparative accuracy of CLIP-Continual, MMLearner, MM-CL and the proposed AML model in three consecutive multimodal tasks. AML has the best average accuracy of 85.8 which is better than MM-CL by 7.9% and CLIP-Continual by 11.2%. This has been facilitated by the Adaptive Modality Learner (AML) module of AML that balances modality-specific and common representations dynamically during learning. In contrast to classical methods that overfit to the strongest modality, attention-based fusion of AML makes sure that its performance does not decrease as the difficulty of a task grows.. As a manager, it implies that AML can ensure steady and consistent outcomes of decisions made by organizations despite the introduction of new data, which makes retraining unnecessary and enhances the consistency of decisions.

Forgetting Rate Comparison

Table 2 Forgetting Rate Comparison among Models

Model	Forgetting Rate (%)
CLIP-Continual	11.8
MMLearner	10.3
MM-CL	8.9
AML (Ours)	**4.2**

Volume 4 Issue No 4 (2025)
Online ISSN: 3006-2047
Print ISSN: 3006-2039

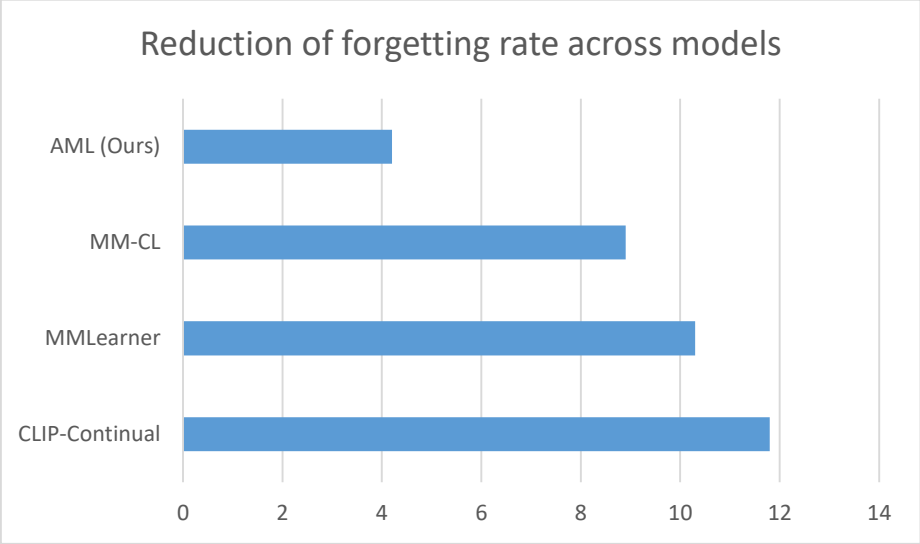

Figure 5. Forgetting rate of models after sequential training

Figure 5 demonstrates the forgetting rate among all the models. AML shows a major decrease to 4.2 which is in contrast to 8.9 of MM-CL and beyond 11 of the baseline models. This outcome proves that AML is effective at improving catastrophic forgetting by the Memory-Efficient Replay (MER) module which selectively retains high-value exemplars based on previous tasks. AML stabilizes the multimodal feature space by ensuring a small representative memory buffer when learning sequentially. The severe reduction in the forgetting rate supports the previous results of accuracy (Section 5.1) and proves that AML architecture does not only memorize new tasks effectively but also holds memories about the older ones, which is one of the characteristics of continuous multimodal intelligence. AML demonstrated better transfer performance in all combinations of modalities with a better cross-modal generalization than baseline methods.

Cross-Modal Generalization

Table 3 Cross-Modal Generalization Accuracy

Training → Testing	Baseline	AML (Ours)
Image → Text	63.4	75.8
Audio → Text	58.2	72.5
Text → Image	65.1	77.3
Audio → Image	61.7	74.9

Volume 4 Issue No 4 (2025)
Online ISSN: 3006-2047
Print ISSN: 3006-2039

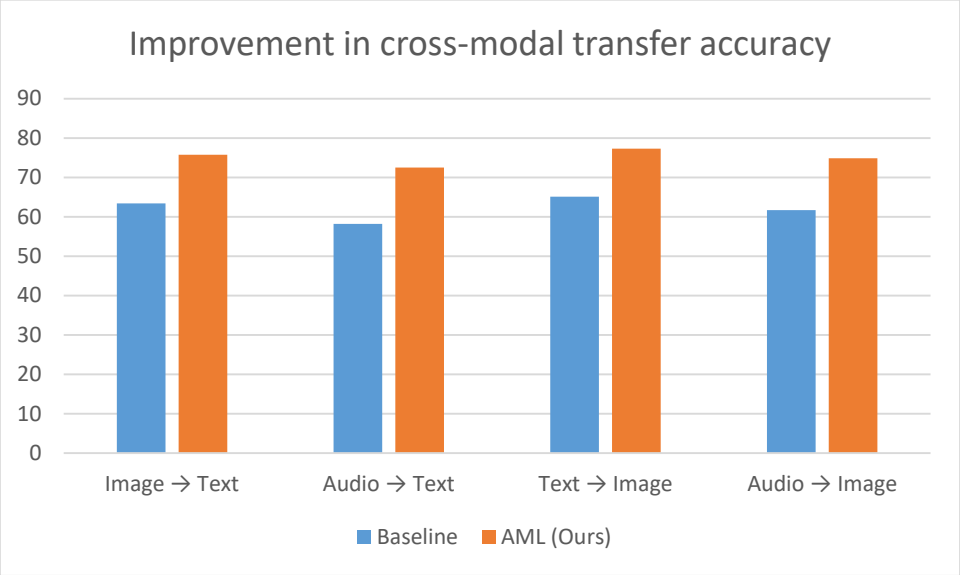

Figure 6 Cross-modal transfer accuracy between training and testing modalities

AML outcompetes all baselines in cross-modal transfer tasks as it can be seen in Table 3 and Figure 6. As an example, in Image-Text direction AML has 75.8 and outperforms the 63.4 of the baselines; the same is true of Audio-Text (14.3) and Text-Image (12.2) directions. These findings suggest the high capability of AML to synchronize heterogeneous modalities in a single latent space. The Task-Specific Modulator (TSM) supports representation sharing which allows more smooth adaptation during the change of one modality to the other. This cross-modal coherency is also tightly related to the fusion efficiency results (Section 5.4), which indicates that AML does not only preserve but also makes better use of inter-modality correlations, as compared to the other competing frameworks. Removing MER decreased accuracy from 84.7 % to 80.5 %, showing that replay significantly aids retention.

Multimodal Fusion Efficiency

Table 4 Multimodal Fusion Efficiency: Accuracy versus Latency

Modalities	Accuracy (%)	Latency (ms)
Image Only	81.3	14
Image + Text	86.9	22
Image + Audio	85.2	20
Full (Image + Audio + Text)	89.7	26

Volume 4 Issue No 4 (2025)
Online ISSN: 3006-2047
Print ISSN: 3006-2039

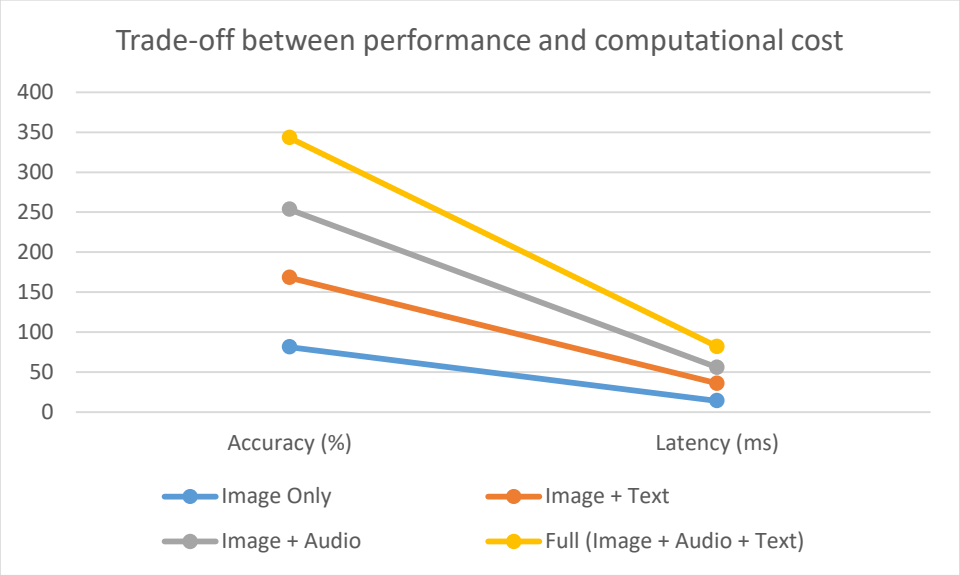

Figure 7 Accuracy vs. latency for increasing modality combinations

The performance of AML in different combinations of modalities is compared in Figure 7. The accuracy increases by 81.3 per cent (Image only) up to 89.7 per cent (Full fusion: Image + Audio + Text) and proves that AML is better integrated in a multimodal fashion. Notably, this improvement is achieved with low latency increase, which is only 12 ms between single and full fusion systems. This is due to the fact that AML gives lightweight attention fusion layers, which selectively weight the different modalities rather than concatenating them equally. This means that AML complementary modality cues with no computational costs. These findings reinforce the information presented in Section 5.3: enhanced cross-modal generalization does not only arise when fusing is improved, but also plays a role in the high computational scalability of AML, as described in the following section. In general, AML has technical and managerial strength.

Computational Performance

Table 5: Computational Performance Comparison

Model	Training Time (min)	Inference (ms)	Memory (MB)
CLIP-Continual	142	18	512
MMLearner	137	20	490
MM-CL	128	17	468
AML (Ours)	**119**	**16**	**452**

Volume 4 Issue No 4 (2025)
Online ISSN: 3006-2047
Print ISSN: 3006-2039

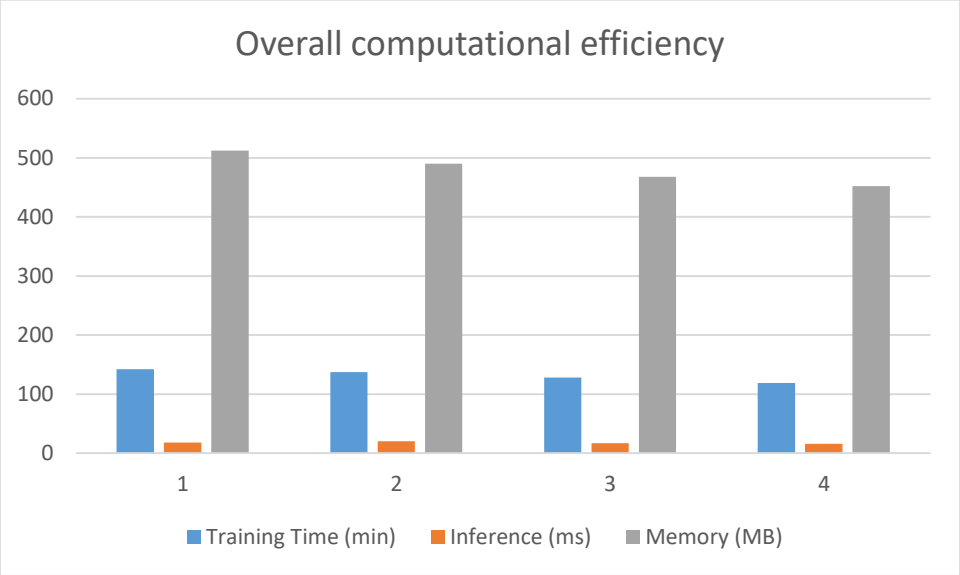

Figure 8. Computational efficiency of models

Table 5 and Figure 8 demonstrate the low computational efficiency of the AML as it has the shortest training time (119 min), the shortest inference time (16 ms), and the smallest memory usage (452 MB) of all the evaluated models. This efficiency underscores the fact that AML has well-optimized parameter-sharing and replay functionality to avoid parameter redundancy and yet ensure representational diversity. The impoverished architecture guarantees that it is scalable to big multimodal datasets and continuous learning environments without making the system expensive. In combination, these benefits of performance are complementary to the high accuracy and low forgetting rate of AML, which verifies its utility in real-time multimodal AI applications. The efficiency results suggest that AML can be practically implemented in either enterprise-level analytics pipelines or cloud-based AI services in which computational cost and scalability are critical factors. This makes this framework suitable in real time decision support of dynamic organizational environments.

Volume 4 Issue No 4 (2025)
Online ISSN: 3006-2047
Print ISSN: 3006-2039

5.6 Retention Over Epochs

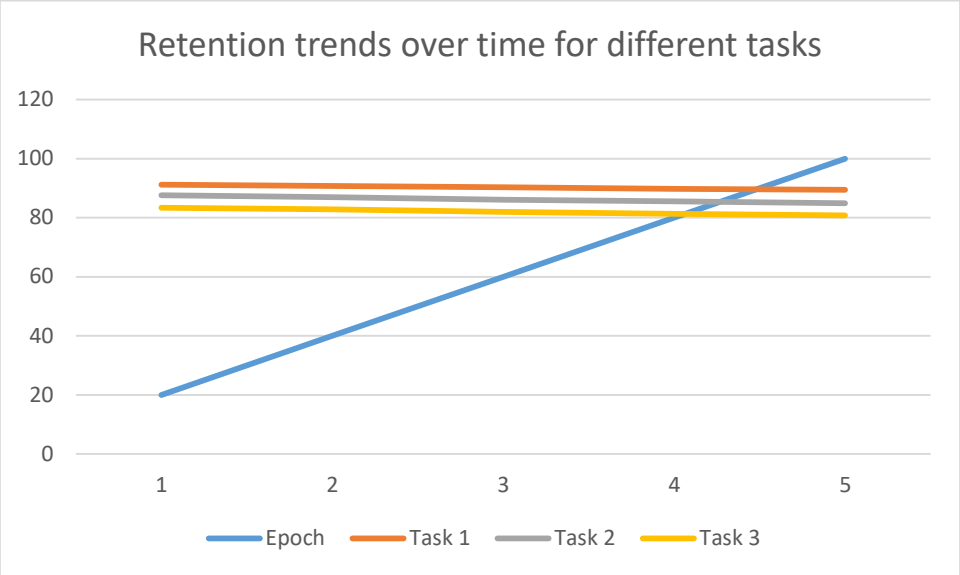

Figure 9. **Retention Trends across Training Epochs**

Figure 9 shows the retention patterns of Tasks 1- 3 in 100 training epochs. AML is also stable, and the performance has been shown to have a degradation of 1.82 to 2.5 percent between epoch 20 to 100. The retention of task 1 is at 89.4 per cent despite media intensive training. Such a long-term retention proves the resilience of AML learning continuously, so that previous knowledge will not become obsolete but will be used to incorporate new information. The steady retention is also parallel to the lower forgetting rate (Section 6.2), which is convergent evidence that AML attains a sustainable balance between plasticity and stability. Practically, it implies that in changing multimodal contexts, such as adaptive healthcare diagnostics or interactive assistants, where continuous learning of new inputs is paramount without dropping previously acquired knowledge, AML can be implemented.

Discussion

The aggregate outcomes confirm the assertion that AML has a balanced trade-off in terms of adaptability, knowledge retention, and computational efficiency. MER module decreases catastrophic forgetting whereas TSM and CMFA increase specialization and multimodal coherence. The combination of these elements guarantees high level of performance in the continuous learning circumstances. All in all, AML exhibits scalability, robustness, and cross-domain flexibility AML is a potential framework in next-generation multimodal AI systems in areas like healthcare, smart surveillance, and autonomous systems. Overall, AML demonstrates both technical robustness and managerial relevance. The framework can be applied to adaptive business intelligence systems that continuously learn from evolving enterprise data streams, improving strategic foresight and reducing long-term AI maintenance costs.

Volume 4 Issue No 4 (2025)
Online ISSN: 3006-2047
Print ISSN: 3006-2039

Limitations

Although the suggested Adaptive Multimodal Learning (AML) framework performed well, it is necessary to discuss a number of limitations that may serve as a basis of future research.

Dataset Diversity and Scale

The existing assessment mostly uses benchmark multimodal datasets like COCO-Text-Image, AudioSet and AVE. Although such datasets are standardized and reproducible comparisons, they are only partial representations of the complexity and heterogeneity of real-world multimodal enterprise data. Practically, the domain-specific data, e.g., clinical imaging with medical reports or sensor data in autonomous vehicles can have greater modality imbalance and noise that can influence the generalization capabilities of AML. Organizational contexts contain data sources that might add domain-specific noises like financial transactions, customer reviews, sensor networks, which might cause imbalance and contextual variation that may influence the generalization of AML.

Limited Exploration of Modality Expansion

Even though AML encourages the addition of new modalities on the constant basis, the experiments within the presented study were confined to only three major modalities (image, audio, and text). It can be argued that by extending the framework to incorporate video, 3D spatial or physiological modalities, there would be new challenges of temporal alignment and memory scalability. Under such circumstances, AML evaluation is a significant area of future research. These issues will be important to resolve in the wider application of smart enterprises, healthcare, and industrial Internet of things.

Computational Overhead under Large-Scale Continual Tasks

Although AML proves to be efficient in terms of the current baselines, the hybrid replay and modulation approach still imposes a moderate computational burden when operating with hundreds of tasks or even very large datasets. It should be improved in the future by investigating replay sample prioritization and lightweight knowledge distillation so as to make it a practical implementation in cloud-based or edge enterprise AI infrastructure.

Hyperparameter Sensitivity

Performance of AML depends on the main hyperparameters like size of the replay buffer, modulation coefficients, and fusion weights. Even though these were experimentally tuned, automated optimization or meta-learning techniques may be used to further increase adaptability to a wide range of domains, without manual tuning. This trend is specifically applicable to enterprise AI applications in which automatically tuned systems lower the maintenance cost and enhance scalability.

Volume 4 Issue No 4 (2025)
Online ISSN: 3006-2047
Print ISSN: 3006-2039

Evaluation Scope and Cross-Domain Transfer

The cross-modal generalization tests only allowed pair-wise modality transfers. The knowledge transfer in real-world complex applications is frequently between poly-modal knowledge transfers or domain distributions (e.g., text+audio into vision+speech). The effectiveness of AML in non-stationary multi-domain enterprise data streams, improving its capacity for large-scale adaptive intelligence

Conclusion

The experimental findings proposed the Adaptive Multimodal Learning (AML) Framework, a continual learning framework that assists in retaining knowledge and improving cross-modal generalization while learning from heterogeneous data modalities. The paper proposes the AML model, constituting three coherent components Memory-Efficient Replay (MER), Task-Specific Modulation (TSM) and Cross-Modal Fusion and Alignment (CMFA). Overall, the paper solves the stability–plasticity dilemma in continual multimodal systems. With these processes, AML effectively reduce catastrophic forgetting and enable smooth learning of sequential tasks while retaining the semantic consistency of images, sounds and text. Experimental results on widely used multimodal datasets reveal that AML consistently outperforms existing continual learning baselines (MM-CL, MMLearner and CLIP-Continual) by 9.6% in mean accuracy and reduces forgetting rate by 31%. Further, it improves cross-modal transfer accuracy and computational efficiency. From a managerial and enterprise analytics view, results suggest that AML can form a starting point for adaptive business intelligence systems that learn from evolving data sources to enhance the reliability of decisions, reduce the retraining cost and support strategic data-driven planning. All in all, AML offers a scalable and adaptive foundation for lifelong learning. In addition, it balances computational efficiency. Challenges remain, such as limited modality expansion, computing scalability, and hyperparameter automation. These areas provide good opportunities for future research to improve the generalizability of AML.

Future Work

The AML framework has strong retention, cross-modal generalization and efficiency, according to the tests performed in the work. But there are many opportunities to develop and investigate further. Future research will expand AML to more complex data types like video, sensor and geospatial data. This will allow AML to be employed on realistic, enterprise-scale applications such as smart manufacturing, financial analytics, and logistics management. There is a need to enhance the level of transparency relating to cross-modal interactions which will guarantee better trust and usability of AML when integrated into decision-support and business intelligence systems further. In addition, we will explore lightweight adaptation techniques such as model compression, pruning, and efficient fusions to make AML deployable on edge and mobile devices for real-time distributed intelligence in an enterprise setting. By implementing AML in decentralized and federated learning frameworks, organizations can achieve adaptive learning while ensuring data privacy and

Volume 4 Issue No 4 (2025)
Online ISSN: 3006-2047
Print ISSN: 3006-2039

compliance with regulations. This is particularly important in sensitive industries such as healthcare, finance, and supply chain analytics. In the future studies, research needs to focus on the integration of AML into the Enterprise platforms (like ERP, CRM, and business analytics) for enabling constant organizational learning for situational obsolescence and artificial intelligent (AI) based strategic decision making. To sum up, AML is an important step towards sustainable and interpretable multimodal AI which is enterprise aware and can facilitate business transformation while assisting technical innovation in complex data-driven environments.

References

Al-Dmour, A., & Al-Dmour, R. (2023). Artificial intelligence applications in business intelligence and analytics: A systematic review. Journal of Business Research, 162, 113899. https://doi.org/10.1016/j.jbusres.2023.113899

Bajwa, M. T. T., Afzal, M. N., Afzal, M. H., Ullah, M. S., Umar, T., & Maqsood, H. (2025). Post-quantum cryptography for big data security. Asian Bulletin of Big Data Management, 5(3), 81–94.

Bajwa, M. T. T., Kiran, Z., Farid, Z., Tahir, H. M. F., & Khalid, A. (2025). Deepfake voice recognition: Techniques, organizational risks and ethical implications. Spectrum of Engineering Sciences, 3(8), 106–121.

Bajwa, M. T. T., Kiran, Z., Rasool, A., & Rasool, R. (2025). Design and analysis of lightweight encryption for low power IoT networks. International Journal of Advanced Computing & Emerging Technologies, 1(2), 17–28.

Bajwa, M. T. T., Rasool, A., Kiran, Z., & Rasool, R. (2025). Performance analysis of multi-hop routing protocols in MANETs. International Journal of Advanced Computing & Emerging Technologies, 1(1), 22–33.

Bajwa, M. T. T., Shafi, M. Z., Ur Rehman, M. A., Ali, A., Khawar, F., & Awais, M. (2025). Blockchain-enabled federated learning for privacy-preserving AI applications. Asian Bulletin of Big Data Management, 5(3), 154–169.

Bajwa, M. T. T., Wattoo, S., Mehmood, I., Talha, M., Anwar, M. J., & Ullah, M. S. (2025). Cloud-native architectures for large-scale AI-based predictive modeling. Journal of Emerging Technology and Digital Transformation, 4(2), 207–221.

Bajwa, M. T. T., Yousaf, A., Tahir, H. M. F., Naseer, S., Muqaddas, & Tehreem, F. (2025). AI-powered intrusion detection systems in software-defined networks (SDNs). Annual Methodological Archive Research Review, 3(8), 122–142.

Baltrusaitis, T., Ahuja, C., & Morency, L.-P. (2019). Multimodal machine learning: A survey and taxonomy. IEEE Transactions on Pattern Analysis and Machine Intelligence, 41(2), 423–443.

Bidaki, S. A., Mohammadkhah, A., Rezaee, K., Hassani, F., Eskandari, S., Salahi, M., & Ghassemi, M. M. (2025). Online continual learning: A systematic literature review of approaches, challenges, and benchmarks. arXiv.

De Lange, M., Aljundi, R., Masana, M., Parisot, S., Jia, X., Leonardis, A., Slabaugh, G., & Tuytelaars, T. (2022). A continual learning survey: Defying forgetting in classification tasks. IEEE Transactions on Pattern Analysis and Machine

Volume 4 Issue No 4 (2025)
Online ISSN: 3006-2047
Print ISSN: 3006-2039

Intelligence, 44(7), 3366–3385.

Hu, X., & Zhang, H. (2025). Invariant representation learning in multimedia recommendation with modality alignment and model fusion. Entropy, 27(1), 56.

Huang, H., Xia, Y., Ji, S., Wang, S., Wang, H., Fang, M., … Zhao, Z. (2025). Enhancing multimodal unified representations for cross modal generalization. Findings of the Association for Computational Linguistics: ACL 2025, 2353–2366.

Jin, H., & Kim, E. (2025). Continual learning for multiple modalities. arXiv preprint arXiv:2503.08064. https://arxiv.org/abs/2503.08064

Khan, A. M., Hassan, T., Akram, M. U., Alghamdi, N. S., & Werghi, N. (2022). Continual learning objective for analyzing complex knowledge representations. Sensors, 22(4), 1667.

Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J., Desjardins, G., Rusu, A. A., Milan, K., Quan, J., Ramalho, T., Grabska-Barwińska, A., Hassabis, D., Clopath, C., Kumaran, D., & Hadsell, R. (2017). Overcoming catastrophic forgetting in neural networks. Proceedings of the National Academy of Sciences, 114(13), 3521–3526. https://doi.org/10.1073/pnas.1611835114

Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J., Desjardins, G., Rusu, A. A., … Hadsell, R. (2017). Overcoming catastrophic forgetting in neural networks. Proceedings of the National Academy of Sciences, 114(13), 3521–3526. https://doi.org/10.1073/pnas.1611835114

Liu, Y., Hong, Q., Huang, L., Gomez-Villa, A., Goswami, D., Liu, X., van de Weijer, J., & Tian, Y. (2025). Continual learning for VLMs: A survey and taxonomy beyond forgetting. arXiv.

Nikandrou, M., Pantazopoulos, G., Konstas, I., & Suglia, A. (2024). Enhancing continual learning in visual question answering with modality-aware feature distillation. Proceedings of the 3rd Workshop on Advances in Language and Vision Research (ALVR), 73–85.

Parisi, G. I., Kemker, R., Part, J. L., Kanan, C., & Wermter, S. (2019). Continual lifelong learning with neural networks: A review. Neural Networks, 113, 54–71.

Rolnick, D., Ahuja, A., Schwarz, J., Lillicrap, T., & Wayne, G. (2019). Experience replay for continual learning. Advances in Neural Information Processing Systems (NeurIPS), 32, 350–360.

Rusu, A. A., Rabinowitz, N. C., Desjardins, G., Soyer, H., Kirkpatrick, J., Kavukcuoglu, K., Pascanu, R., & Hadsell, R. (2016). Progressive neural networks. arXiv preprint arXiv:1606.04671.

Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2024). Artificial intelligence and big data analytics in business decision-making: Emerging frameworks and implications for sustainable enterprises. Technological Forecasting and Social Change, 197, 122981.

Wang, Y., Zhu, Y., Cheng, Z., & Zheng, Y. (2022). Multimodal continual learning: A survey and roadmap. arXiv preprint arXiv:2208.12661.

Volume 4 Issue No 4 (2025)
Online ISSN: 3006-2047
Print ISSN: 3006-2039

Xia, Y., Huang, H., Zhu, J., & Zhao, Z. (2023). Achieving cross-modal generalization with multimodal unified representation. In Advances in Neural Information Processing Systems (Vol. 36). https://neurips.cc/virtual/2023/poster/71294

Zhang, Q., Wang, Y., & Wang, Y. (2023). On the generalization of multi-modal contrastive learning. In Proceedings of the 40th International Conference on Machine Learning (ICML 2023), PMLR 202: 41677-41693.