STM JOURNALS

Recent Trends in
Programming Languages

ISSN: 2455-1821
Volume 12, Issue 3, 2025
September–December
DOI (Journal): 10.37591/RTPL

https://journals.stmjournals.com/rtpl

Review RTPL

Continuous Learning in Language Models: A Survey of Streaming Data Processing Techniques

Hemant N. Patel*

Abstract

The integration of continual learning with Large Language Models (LLMs) and Natural Language Processing (NLP) represents a transformative step toward creating adaptive, intelligent systems capable of functioning effectively in ever-changing environments. Traditional LLMs are typically trained on large, pre-collected datasets, which limits their ability to evolve as new information emerges. Continual learning, in contrast, enables models to acquire new knowledge incrementally without the need for complete retraining, thereby supporting long-term adaptability and efficiency. This study explores the theoretical foundations of lifelong learning from both human cognitive science and machine learning perspectives, highlighting parallels between human neuroplasticity and artificial adaptability. It also examines advanced NLP techniques for real-time text processing, data preprocessing, and contextual understanding that enhance dynamic system performance. Furthermore, the discussion extends to the role of NLP in data visualization, streaming text analytics, and semantic feature extraction, demonstrating its synergy with continual learning frameworks. Collectively, this synthesis offers a holistic overview of existing methodologies and establishes a conceptual foundation for developing more responsive, intelligent, and sustainable Artificial Intelligence (AI) systems capable of continuous evolution.

Keywords: Continual learning, lifelong learning, large language models, natural language processing, streaming data, real-time text processing, data preprocessing, neuroplasticity, information literacy, adaptive AI systems

INTRODUCTION

There is a great deal of promise for general intelligence algorithms in large language models. As the scale of parameter size increases, researchers have shown that complicated skills including instruction adhering to, few-shot in-context learning, and multi-step logic get better [1]. The significance and innovative nature of LLM progress has caused machine learning experts to reevaluate conventional computing approaches for tasks that were previously difficult for humans [2]. But because LLMs are usually trained on static, pre-gathered datasets covering generic domains, their performance gradually deteriorates over time and across various content domains. Furthermore, a single huge model that has already been trained cannot satisfy all user needs and needs to be continually refined.

***Author for Correspondence**
Hemant N. Patel
E-mail: hp15284@gmail.com

Assistant Professor, Department of Computer Engineering, Sankalchand Patel College of Engineering, Sankalchand Patel University, Visnagar, Gujarat, India

Received Date: July 04, 2025
Accepted Date: September 30, 2025
Published Date: October 15, 2025

Citation: Hemant N. Patel. Continuous Learning in Language Models: A Survey of Streaming Data Processing Techniques. Recent Trends in Programming Languages. 2025; 12(3): 23–34p.

Re-collecting initial training data and re-training models with new specific demands is one possible option, however this method is unfeasible and unworkable in realistic situations.

The idea of lifelong learning is not new. Over a long time, it has been defined in a number of different ways. In spite of the expansion of the "knowledge" society and the numerous societal, financial, and academic shifts it seems to imply,

international organizations like the OECD, UNESCO, and the European Union are presently aggressively promoting the concept of continuous education [3]. Many national governments in Europe have taken up the issue and made it a far more significant component of their political plans [4]. A growing percentage of knowledge, especially advanced knowledge, is acquired via non-traditional methods of learning, frequently far beyond the formal schooling age.

The goal for ongoing learning is to develop machine learning techniques that can learn tasks given as an unpredictable stream. In general, CL's reduced computing costs as compared to offline retraining on all accessible data demonstrate its utility for real-world machine learning systems. Although offline retraining may solve CL trivially in an imaginary unlimited compute regime, experience replay, a technique that uses buffers of few samples per task is employed in practice [5]. Compared to other families of CL approaches, this approach is acknowledged to be highly adaptable and suitable in a wide range of settings. The establishment of class progressive learning and ER as the preferred approach are the main topics of this study.

Natural Language Processing is a branch of Artificial Intelligence and Linguistics aimed at enabling computers to understand and interact using human language. It helps users communicate with machines without needing specialized programming knowledge [6]. NLP involves two main components: Natural Language Understanding and Natural Language Generation, which focus on interpreting and producing human language, respectively (Figure 1).

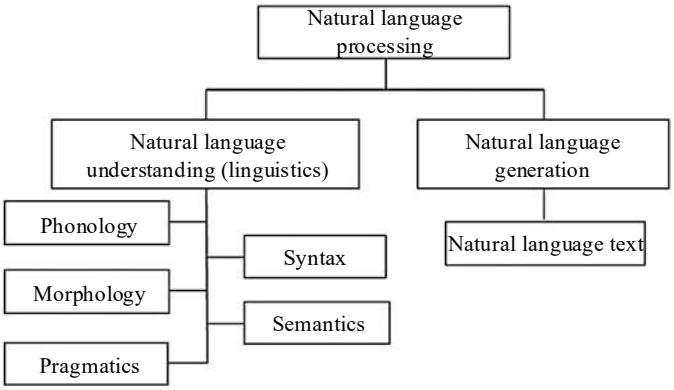

Figure 1. Broad classification of NLP.

The Study's Structure

This study's structure is set up to give a thorough rundown of important subjects in language models and continuous learning. It begins with the fundamentals of Continuous Learning, followed by a discussion on Streaming Data in Natural Language Processing. Then it explores Architectures for Continuous Learning in Language Models, followed by Techniques for Streaming Data Processing. Then it presents a detailed review of recent literature. Finally, the last Section concludes the study with insights into future research directions.

FUNDAMENTALS OF CONTINUOUS LEARNING

Numerous instructional foundations, which have also gradually emerged with the advancements in lifetime learning, are essential to its success. "Adult learning" is one of the oldest and most hospitable educational programs. Actually, it is aimed for third-age students and focuses more on basic literacy, that would enable people to pursue further education [7]. The goal of bibliographic teaching is to equip people to use knowledge, data resources, and information systems effectively throughout their lives, in addition to giving them the specialized skills they need to finish tasks. Compared to any other educational pathway, knowledge of information, computer and internet literacy, and lifelong learning have been integrated together in recent years [8]. As a necessary component for people to be educated in the field, information literacy has now permeated the whole information society. A solid foundation

of knowledge and literacy in information and communication technology is necessary for effective information searching and continuous learning.

Continuous Learning

In the contemporary technological age, the concept of continuing education or continuous learning places an emphasis on understanding, skills, and abilities related to information and technology advancements [9]. In addition, it makes the case that adult education is essential for giving people the chance to actively engage in all facets of life, irrespective of their age, gender, social standing, or economic standing. It is implicitly assumed that this is done in order to build up beneficial traits and competencies through ongoing high-quality education that complies with the demands of the global age. The high-quality education and learning are essential in the age of globalization because they are seen to be able to produce people with superior human qualities including creativity, innovation, productivity, skill, competitiveness, and challenge. According to Cardon, workers must constantly improve and update their knowledge and skills; as a result, lifelong learning will occur throughout a person's whole life. The education in institutions, businesses, and groups in the next years will probably result in instructor companies which can foresee shifts and variety in human resources' knowledge, skills, and capacities, improving their effectiveness.

Types of Learning Paradigms

Modern AI systems require adaptive capabilities to deal with dynamically evolving data. Three prominent paradigms that support such adaptability are Incremental Learning, Online Learning, and Lifelong Learning. These paradigms differ in how they process data over time and manage knowledge accumulation.

Incremental Learning

Incremental Learning refers to the paradigm where the model is trained continuously with new data batches without retraining from scratch. It retains previously learned knowledge while integrating new information, thereby mitigating catastrophic forgetting. Incremental approaches are particularly effective for scenarios with periodically arriving data or concept drift.

Online Learning

Online Learning is designed for real-time data processing, where the model updates its parameters immediately after each new data point or small batch. Unlike incremental learning, which may wait for a data chunk, online learning emphasizes responsiveness and is suitable for streaming environments. It is ideal for systems requiring low-latency learning and decision-making.

Lifelong Learning

The capacity to consistently pick up new information and abilities throughout one's life is known as ongoing education in the human brain. The brain's neuroplasticity, the capacity to forge new neural connections and fortify pre-existing ones, makes it feasible [10]. Continuous education, which enables us to adjust to changing circumstances and difficulties, maintain mental acuity and engagement, try new things, acquire new skills and information, and make significant contributions to society, depends on both explicit and implicit learning. Two of the most significant aspects influencing the human brain's lifelong learning process have been studied and are listed as follows:

- *The Biological Theory of Neuroplasticity:* The biological theory of learning and neuroplasticity states that neurons have dendrites for receiving information and axons for sending it out. When two neurons' axons are close to each other, they can communicate through a synapse; with repeated activation, a neuron's dendrites can thicken and lengthen, enhancing its capacity for interaction with adjacent neurons for exchange of data.
- *The Biology of Forgetting:* It is clear from everyday experience that if knowledge is not promptly revised, it tends to be forgotten over time. In order for the brain to give priority to relevant data and eliminate unnecessary information, forgetting is a normal occurrence.

- *Quantifying Forgetting:* The Ebbinghaus Forgetting Curve, which may be seen as an exponential decay curve, explains how knowledge is lost over time.

Continual Learning for LLMs

The goal of every day Learning for Large Language Models is to allow LLMs to continuously learn from a stream of data throughout time. Despite its significance, applying current continuous learning environments right away to LLMs is not simple [11]. A classification of studies in this field and offered a forward-looking framework of continuous learning for LLMs. Figure 2 shows the framework of continuous learning for LLMs.

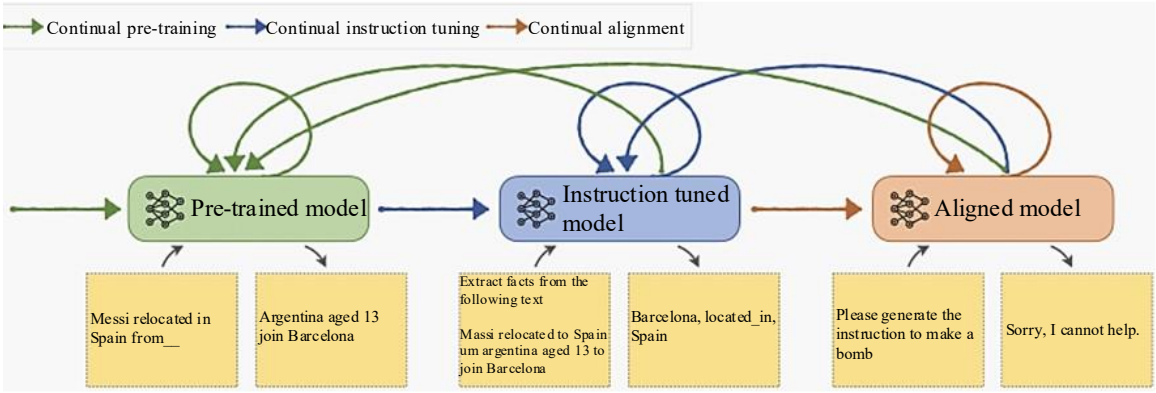

Figure 2. Continual learning in LLMs.

Align LLMs' ongoing education with the various training phases, such as CPT, CIT, and CA. The goal of the Continuous Pre-training stage is to train LLMs on a series of self-supervised corpora in order to expand their expertise and help them adapt to new fields. The ongoing education using a stream of supervised instruction, following data, the tuning step optimizes LLMs [12], with the goal of enabling LLMs to carry out human commands while applying learned information to future tasks. Over time, CA works to continually align LLMs with human values in response to the changing nature of human values and choices.

STREAMING DATA IN NATURAL LANGUAGE PROCESSING (NLP)

Using big data processing and the techniques described in this section, the goal is to set up a linguistic pipeline that can automatically extract or find knowledge in vast volumes of texts. Textual data must be processed in parallel for scalable NLP processing [13]. Effective parallelization can be carried out at several levels, ranging from reimplementing each module's basic algorithms utilizing multi-threading, parallel computing to distributing copies of the same LP among servers [14]. Since it is impractical to re-implement every module required to carry out a task as complicated as mining events, this final kind of fine-grained parallelization is obviously outside the purview of the current study. By developing and putting into practice an NLP architecture that enables the simultaneous processing of texts, it seeks to handle enormous amounts of textual data.

NLP in Data Visualization

NLP is essential for deciphering unstructured data, which makes up a large amount of data produced in the digital world. Because unstructured data, including text from emails, reports, and social media, does not have a predetermined data model, it might be difficult to evaluate using conventional techniques [15]. To find useful information from this data, NLP techniques including word mining, sentiment analysis, and NER are used. Text mining is the process of examining vast volumes of text to find correlations, patterns, and trends that make it easier to find pertinent information [16]. Another significant method is sentiment analysis, which evaluates the feelings, viewpoints, and attitudes conveyed in text to reveal customer preferences and public mood. NER helps to extract structured data,

such as names that can be grouped into types (person, organizations, locations), out of unstructured data very often by identifying and classifying Named Entities in a document or other text.

Challenges of Processing Streaming Text

Real-time processing of streaming text leads to a few complications that are not characteristic of traditional batch processing [17]. These difficulties are caused by the non-structured, evolving, and dynamic nature of the information that requires adaptive and efficient learning systems. The major challenges in this field can be seen below:

Real-Time Processing Requirements

Text messages are also received in a streaming manner and have to be handled at nearly the same rate as they are received to ensure stream responsiveness of the systems. This puts stringent limits on latency of language models and data pipelines [18]. Slow responses are a hindrance to the real-time applications like making traffic, analysis of finances, or a response to emergency. It is of great significance to be fast and accurate.

Concept Drift and Data Evolution

The behavior and pattern of data in streaming settings may evolve with time a phenomenon called concept drift. A language model that was studied on data that is older might be unable to understand new trends and topics as well as slang. In the absence of adaptive mechanisms, performance of the models reduces. The active development makes this a process that needs constant model updates or optimization.

Noisy and Unstructured Data

Sources of streaming text such as social media or chats may present informal language, spelling mistakes, abbreviations and useless information. Such quality problems make correct parsing and analysis hard. The noise filtering pre-processing methods have to be lightweight and efficient as to not create latency.

Catastrophic Forgetting

When language models are updated incrementally with new data, they risk overwriting or forgetting previously acquired knowledge. This is known as catastrophic forgetting. Designing architectures that retain long-term knowledge while learning from streaming inputs remains an open research challenge in continual learning.

Scalability and Resource Constraints

Processing large-scale, high-velocity text streams requires scalable infrastructure. However, resource limitations especially in memory and compute make it difficult to store past data or train models continuously. Efficient model compression, sampling, and memory management strategies are needed for deployment in real-world systems.

ARCHITECTURES FOR CONTINUOUS LEARNING IN LANGUAGE MODELS

Streaming data

Continuous learning architectures of language models are intended to overcome the shortcomings of traditional, so-called static models, which must be retrained on entire datasets in order to fit to new data. This is surmounted by continuous learning models that include features to enable them learn over time in real time without catastrophic forgetting. Such architectures usually encompass rehearsal-based approaches that keep a backlog of previous data to periodically re-train the network; regularization-based approaches such as Elastic Weight Consolidation that save significant weights; and dynamical arch-techniques that scale or re-model the model when new tasks arise [19]. The other models also make use of external memory modules to run episodic recall and meta-learning methods to adapt quickly based on less information. Such advances inspire the language models to process streaming information efficiently in dynamic conditions. Traditional vs. continuous learning language models' comparison is shown in Table 1.

Table 1. Modern continuous learning architecture vs. traditional language model architectures.

Aspect	Traditional Language Models (BERT, GPT, etc.)	Modern Continuous Learning Architectures
Training Paradigm	One-time offline training on a fixed corpus	Continuous/incremental learning on streaming or evolving data
Adaptability to New Data	Requires full retraining or fine-tuning	Supports online updates or incremental learning
Handling of Catastrophic Forgetting	Prone to forgetting past knowledge during fine-tuning	Incorporates mechanisms to preserve prior knowledge (e.g., memory replay, EWC)
Vocabulary Expansion	Fixed vocabulary and embeddings	Dynamic vocabulary and embedding updates
Model Update Cost	Computationally expensive and time-consuming	Designed for lightweight, frequent updates
Memory Utilization	No explicit memory of past inputs	Uses episodic memory or memory-augmented networks
Architecture Flexibility	Monolithic and static	Modular and extensible (e.g., adapters, progressive modules)
Suitability for Streaming Data	Not suitable; processes batch data only	Suitable for real-time and non-stationary data streams
Example Models	BERT, GPT-2, RoBERTa, XLNet	Online BERT, Continual Transformer, Adapter Fusion, Memory-Augmented Models

Adaptations for Continuous Learning

In order to allow language models to work in non-stationary setting or process streaming data, there are some adaptations suggested. Other techniques like EWC and Replay-Based Training are built up on top of existing architectures to standardize updates and maintain previous knowledge. Simultaneously, the alteration of architectural elements (using embedding layers, memory blocks and attention blocks) has facilitated a more flexible and module-based workaround to lifelong learning.

- *Dynamic Embedding Layers:* These reformat to either new vocabulary or new context or not disturbing the representation of the old ones by simply extending its embedding space. Such advances as vocabulary hashing, incremental updates, and meta-embedding fusion allow constant extension of unobserved or domain-specific terms.
- *Episodic Memory Integration:* Episodic memory modules are used to recall past experience using an external buffer, which ensures that the model recalls previous knowledge and decreases forgetting. MANNs, Lifelong Memory Networks and other architectures combine these memories with transformers as context-aware predictors.
- *Modular architectures:* Modular architectures split learning into independent components that can be revised and re-run as modules. Such models as Progressive Neural Networks and Adapter-based Transformers enable adding new modules to do different tasks without performing another training of the whole network.

Role of Transformers in Online Learning with Applications of Continuous Learning in NLP

The flexible transfers provided by the parallel, layer-wise representations and the attention mechanisms in transformers make them well adapted to on-line learning. Nevertheless, the conventional transformer design should be adjusted to efficiency in real-time or in perpetual learning scenario. Streaming attention, sliding-window context, attention masking and memory-efficient recurrence techniques enable transformers to process streaming inputs, whilst maintaining critical contextual dependencies. Recent work also explores Online Transformers and Efficient Lifelong Transformers that dynamically adjust their computation and storage for sequential updates.

Here are the key applications of Continuous Learning in NLP:

- *Real-Time Sentiment Analysis:* Continuous learning allows sentiment models to adapt to evolving language and opinion trends, ensuring accurate real-time analysis in domains like finance, social media, and customer feedback.

- *Spam and Fake News Detection:* Continual adaptation helps models keep up with new spam tactics and misinformation patterns, maintaining robust performance without full retraining.
- *Conversational Agents and Chatbots:* Chatbots benefit from continuous updates by learning from user interactions, enabling improved personalization and handling of new intents over time.
- *Language Model Adaptation in Low-Resource Domains:* In fields with limited data, such as medicine or law [20], continuous learning helps fine-tune language models with incremental updates, improving performance while avoiding catastrophic forgetting.

TECHNIQUES FOR STREAMING DATA PROCESSING

One of the few highly prevalent professions in data mining is data processing, which involves organizing and transforming data so that it can be mined [21]. Data preparation aims to reduce data size, identify links between data, standardize data, remove outliers, and retrieve data attributes. Among the techniques employed are data reduction, integration, transformation, and purification.

The General Framework for Building Operational Data Preprocessing

Data purging, elimination, scaling, conversion, and segmentation are the five main activities that make up creating operational data preparation in general [22]. By removing outliers and reproduction of values that are missing, data clean-up seeks to improve the quality of the data. Usually, two-dimensional data tables are used to record building operational data, with each column denoting a building variable and each row representing an observation made at a particular time step.

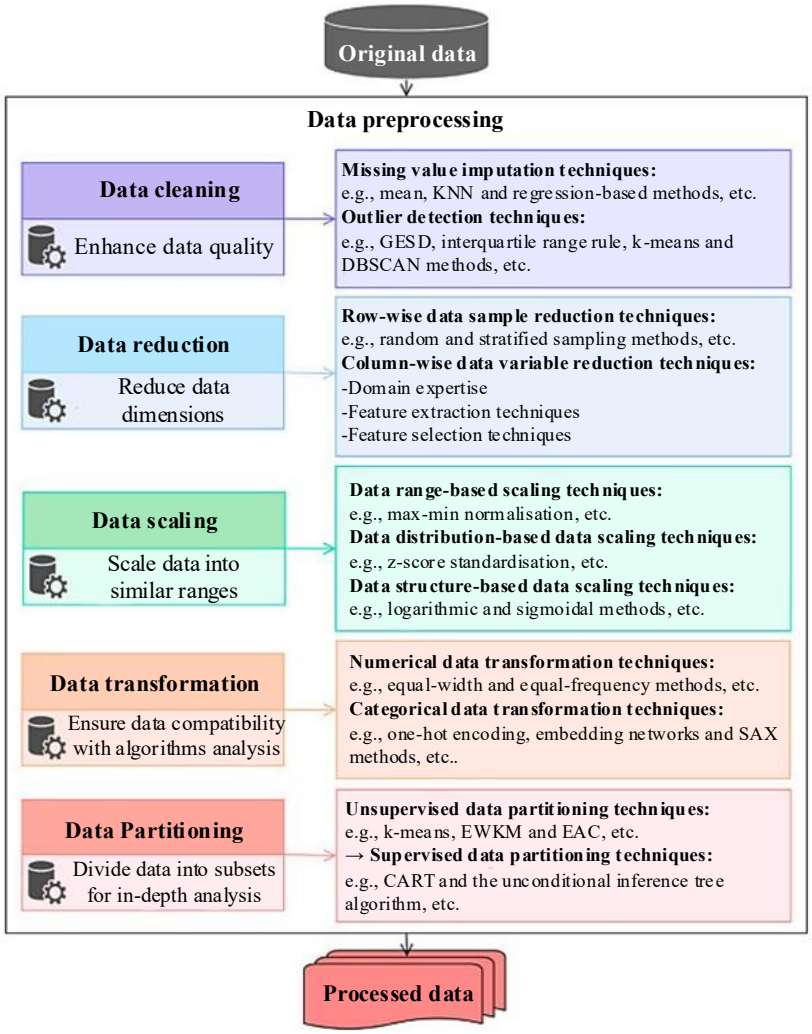

Figure 3. Common preprocessing activities for operational data analysis.

Data reduction in this situation may be done in two ways: column-wise for data variable reduction and row-wise for data sample reduction. Data reduction is used to decrease the dimensionality of data and, consequently, the related processing expenses. Data scaling, which may be accomplished in three primary ways: data range, distribution, and structure-based methods, aims to convert the original data into comparable ranges for predictive modelling as shown in Figure 3. Organizing the original data into forms that are appropriate for different data mining methods is the goal of data conversion. It normally contains two jobs, i.e., numerical information conversion which changes numerical data to categorical data, and categorical data transformation which translates categorical data into numerical data.

Data Preprocessing in Streaming Contexts

Pre-processing of data is a very important ingredient in streaming scenarios where the quality and relevance of the incoming text need to be verified before passing it on to the continuous learning systems. In contrast to static data, streaming text comes in an unbounded flow in real-time, usually with noise, informal text, and contradictions. This requires light weight, fast pre-processing strategies, capable of working at a low latency level [23]. The most notable pre-processing is to tokenize, eliminate stop words, normalization, and even eliminate irrelevant types of filtrations or duplicated content. Also, in the case of streaming systems, they should continually adjust to new language flows, emerging words or even the use of jargon. The problem is that pre-processing pipelines should be designed to be scalable and responsive to avoid the decrease of the quality of the data without impacting the speed of the process necessary to ensure that learning and decision-making is real-time. Also, classic batch pre-processing approaches tend to be too resource-demanding or time-consuming when working with streaming. Hence, real-time systems have to use incremental and memory efficient algorithms. In real world applications, pre-processing should be able to cope with multi-lingual and code-mixed document, is an addition to the complexity.

Feature Extraction and Vectorization in Real-Time

Finding the features and creating a vector is a very important procedure in the extraction and symbolization of raw streaming of texts to be processed by language models [24]. When streaming data is concerned, these tasks must be carried out in a timely manner using computationally lightweight methods without slowing down the data stream that can be processed. Content-Based transformation which includes TF-IDF or Bag-of-Words is based on a fixed vocabulary but each record may contain any number of words, and as examples using listings of spreadsheets breaks some of the constraints of the memory and space usage that the use of feature vectors of TF-IDF entails, even when using sparse matrix formats [25]. The alternatives to content-based transformations to be, attained more expedient forms of obtaining a representation of the text through very computational and dynamic formulations including incremental word embeddings based on the dynamically evolving vocabularies; hashing tricks what is augmented with TF-IDF and others, or on transformer encoders to produce feature vectors representation by using pretrained transformer encoders as in the approach Google suggests called BERT, the memory allocation with loaders in a GPU can be compared relative to a pancake mow down with someone that is penning tunes the computer.

LITERATURE REVIEW

This section presents a concise overview of literature on continuous learning in language models, highlighting the benefits of real-time model updates, personalized learning environments, and adaptive data processing through the integration of NLP, recommendation systems, and dynamic learning frameworks (Table 2).

Kalaiselvi *et al*. (2025) [26]

Large volumes of unmanaged traffic-related data, including social networking updates, traffic reports, and incident logs, may be interpreted and analyzed using NLP. The use of a BERT classifier that allows extracting and classifying relevant traffic data that lies in a textual source in a more accurate way is an essential component of this plan. This data can help the BERT-based NLP system to identify

the pattern of congestion, provide a more precise estimate and optimizing traffic dynamically according to the analysis of this data in real-time. The experiment proves the potential of NLP, particularly as combined with BERT, in enhancing the responsiveness, and productivity of traffic control systems, which will ultimately culminate in the reduction of time spent on transit, emissions, and urban transportation [26].

Li *et al*. (2024) [27]

A thorough examination of typical data processing methods was applied to contemporary multimodal model training, with an emphasis on MLLMs and diffusion models. It categorized all methods into four groups: data safety, data distribution, data quantity, and data quality. The purpose of this study is to give developers of multimodal models advice on efficient data processing methods [27].

Glushkova *et al*. (2024) [28]

Glushkova *et al*. provided a personalized approach to lifelong learning in a school education platform. The considered approach is presented in the BLISS platform, which is an adaptation of the ViPS reference architecture and is prototyped as a cyber-physical and social system. Users interact with the system through their personal assistants. The presented platform can implement personalized training through an adapted LMS; an e-Diary modeled by blockchain technology, as well as a customized Test system [28].

Table 2. Literature of review on continuous learning in language models.

Author	Study On	Method	Key Findings	Challenges	Future Directions
Kalaiselvi *et al*. (2025) [26]	Real-time NLP for traffic data using BERT	BERT-based classifier on streaming traffic data	Improved traffic predictions, congestion detection, real-time insights	Processing noisy and unstructured text data in real-time	Expand NLP use in other smart city applications with continuous adaptation
Li *et al*. (2024) [27]	Data processing in multimodal language model training	Classification into data quality, quantity, distribution, and safety	Summarized effective data processing strategies for MLLMs and diffusion models	Generalizing processing techniques across varied model types	Design adaptive processing pipelines for continuous multimodal learning
Glushkova *et al*. (2024) [28]	Personalized lifelong learning in school LMS	BLISS platform using ViPS architecture and blockchain	Personalized learning via LMS, e-diary, and testing modules	Scalability of personalized models and integration with learning agents	Apply similar frameworks to adaptive AI-based tutoring systems
Murad *et al*. (2023) [29]	Online learning personalization in LMS	Customized recommendation system based on interaction history	Identified key variables influencing adaptive recommendations in MOOCs	Handling large-scale personalized data with privacy	Develop real-time adaptive LMS integrated with continuous user profiling
Xu *et al*. (2021) [30]	Dynamic model updating for mechanical systems	Improved updating of MGDM (multi-stage gearbox)	Accuracy significantly improved in dynamic model prediction	Time-consuming parameter tuning and model complexity	Automate continuous updating in complex real-time systems
Guo *et al*. (2021) [31]	Distributed Streaming Data Processing Systems	Comparative analysis of 3 DDSPS frameworks	High scalability, fault tolerance, and real-time data processing	Real-time reliability and data consistency	Use DDSPS for robust streaming text processing in language models

Murad *et al*. (2023) [29]

Murad *et al*. find ways to improve online learning by employing a personalized recommendation system that is based on the user's past interactions with the system and uses a number of variables or determinants. The approach is a review that is carried out in phases. This leads to an understanding of the user's influencing elements and the opportunities in the system. Learning management systems are presently frequently employed for individualized development suggestions. Both the web and mobile devices may be used to access the massively open online courses. One of the learning resources that is frequently utilized in online education nowadays is the LLMS [29].

Xu *et al*. (2021) [30]

A more accurate forecast of the dynamic properties of the real system was made possible by the revised MGDM's considerable improvement in accuracy when compared to the original model. This demonstrates how well the enhanced updating technique is. The multistage gearbox dynamic model's accuracy may be increased by upgrading the model. However, prior updating techniques are not appropriate for updating the MGDM due of its many parameters and laborious computations. An enhanced dynamic model updating technique appropriate for updating the MGDM has been developed to address the aforementioned issues [30].

Guo *et al*. (2021) [31]

High real-time speed is necessary for data stream processing, while durability and dependability are necessary for data computation. The data stream processing issue in the big data context may be resolved using DDSPS. It provides strong real-time processing capabilities in addition to the benefits of distributed systems' scalability and fault tolerance. Three open source distributed streaming data processing systems are presented in this article along with a comparison and analysis of the three streaming frameworks. Technical references for the theoretical investigation and application technology development of data stream processing in the big data environment [31].

Table 2 gives a summary of the research, methodology, main conclusions, difficulties, and future prospects of the literature on continuous learning in language models.

CONCLUSION AND FUTURE WORK

This review highlights the convergence of continual learning, NLP, and LLMs as a pivotal advancement toward achieving AGI. It underscores the limitations of static training in LLMs and the critical need for adaptive learning frameworks that can process streaming data in real time, overcome catastrophic forgetting, and respond to evolving language patterns and user needs. By integrating insights from neuroscience, data pre-processing strategies, and lifelong learning paradigms, the review presents a compelling case for embracing continual learning methodologies to enhance LLM capabilities across dynamic and unstructured environments. However, current continual learning systems still struggle with scalability in large-scale deployments and lack standardized protocols for evaluating long-term retention. Additionally, maintaining model stability while integrating new knowledge remains a significant challenge.

Future research should focus on developing scalable, low-latency continual learning algorithms tailored for LLMs that can effectively handle concept drift, support multi-lingual and code-mixed inputs, and minimize resource consumption. Innovations in lightweight pre-processing, incremental vectorization, and biologically inspired memory systems will be critical. Furthermore, creating standardized benchmarks and evaluation metrics for continual learning in NLP tasks, particularly within streaming contexts, is essential to facilitate robust model comparison and drive real-world deployment of adaptive, intelligent systems.

REFERENCES

1. Shi H, et al. Continual learning of large language models: A Comprehensive Survey. ACM Comput Surv. 2024; 1(1): 1–44.

2. Khare P, Arora S, Gupta S. Integration of Artificial Intelligence (AI) and Machine Learning (ML) into Product Roadmap Planning. In 2024 IEEE First International Conference on Electronics, Communication and Signal Processing (ICECSP). 2024 Aug 8; 1–6.
3. Alla X. Lifelong learning. Interdiscip J Res Dev. 2024 Mar 23; 11(1): 27–32.
4. Chatterjee P. Real-Time Payment System and their Scalability Challenges. Iconic Res Eng J. 2023; 6(12): 1461–1470.
5. Choudhary P, Jalan V. Enhancing Process Comprehension through Simulation-Based Learning. Int J Adv Res Sci Commun Technol. 2022 Dec; 2(2): 919–24.
6. Khurana D, Koli A, Khatter K, Singh S. Natural language processing: state of the art, current trends and challenges. Multimed Tools Appl. 2023 Jan; 82(3): 3713–44.
7. Asundi AY, Karisiddappa CR. Foundations of lifelong learning and the objective role of LIS Education connoisseurs. In World Library and Information Congress: 72nd IFLA General Conference and Council. 2006 Feb 6.
8. Dattangire R, Vaidya R, Biradar D, Joon A. Exploring the Tangible Impact of Artificial Intelligence and Machine Learning: Bridging the Gap between Hype and Reality. In 2024 IEEE 1st International Conference on Advanced Computing and Emerging Technologies (ACET). 2024 Aug 23; 1–6.
9. Budiningsih I, Soehari TD, Supriyanto E. Continuous learning for employee capacity developing in personal mastery at Bank Indonesia. Indones J Learn Adv Educ. 2022 Dec 16; 5(1): 61–77.
10. Peng J, Sun X, Deng M, Tao C, Tang B, Li W, Wu G, Liu Y, Lin T, Li H. Learning by active forgetting for neural networks. arXiv preprint arXiv:2111.10831. 2021 Nov 21.
11. Zheng J, Qiu S, Shi C, Ma Q. Towards lifelong learning of large language models: A survey. ACM Comput Surv. 2025 Mar 7; 57(8): 1–35.
12. Pandya S. Comparative Analysis of Large Language Models and Traditional Methods for Sentiment Analysis of Tweets Dataset. Int J Innov Sci Res Technol. 2024; 9(12): 1647–57.
13. Agerri R, Artola X, Beloki Z, Rigau G, Soroa A. Big data for Natural Language Processing: A streaming approach. Knowl-Based Syst. 2015 May 1; 79: 36–42.
14. Rathore PS, Sharma BK. Business Intelligence Tools in 2024: A Comparative Analysis and Market Insights. Journal of Global Research in Electronics and Communication (JGREC). 2025 May; 1(5): 18–22.
15. Uddin MK. A review of utilizing natural language processing and AI for advanced data visualization in real-time analytics. International Journal of Management Information Systems and Data Science. 2024 Apr 20; 1(4): 34–49.
16. Rongala S, Pahune SA, Velu H, Mathur S. Leveraging Natural Language Processing and Machine Learning for Consumer Insights from Amazon Product Reviews. In 2025 IEEE 3rd International Conference on Smart Systems for applications in Electrical Sciences (ICSSES). 2025 Mar 21; 1–6.
17. Mehmood E, Anees T. Challenges and solutions for processing real-time big data stream: a systematic literature review. IEEE Access. 2020 Jun 26; 8: 119123–43.
18. Murri S, Bhoyar M, Selvarajan GP, Malaga M. Transforming Decision-Making with Big Data Analytics: Advanced Approaches to Real-Time Insights, Predictive Modeling, and Scalable Data Integration. Int J Commun Netw Inf Secur. 2024; 16(5): 506–19.
19. Choudhary P, Choudhary R, Garaga S. Enhancing training by incorporating ChatGPT in learning modules: an exploration of benefits, challenges, and best practices. Int J Innov Sci Res Technol. 2024; 9(11): 1578–1582.
20. Pahune S, Chandrasekharan M. Several categories of large language models (llms): A short survey. arXiv preprint arXiv:2307.10188. 2023 Jul 5.
21. Saraswat P, Raj S. Data pre-processing techniques in data mining: A Review. Int J Innov Res Comput Sci Technol. 2022; 10(1): 122–125.
22. Fan C, Chen M, Wang X, Wang J, Huang B. A review on data preprocessing techniques toward efficient and reliable knowledge discovery from building operational data. Front Energy Res. 2021 Mar 29; 9: 652801.
23. Ramrez-Gallego S, Krawczyk B, Garca S, Woniak M, Herrera F. A survey on data preprocessing for data stream mining. Neurocomputing. 2017 May 24; 239(C): 39–57.

24. Bathla G, Singh P, Singh RK, Cambria E, Tiwari R. Intelligent fake reviews detection based on aspect extraction and analysis using deep learning. Neural Comput Appl. 2022 Nov; 34(22): 20213–29.
25. Chatterjee P, Das A. Leveraging Machine Learning for Predictive Bug Analysis. Int J Sci Res Manag. 2024 Dec 16; 12(12): 1804–1814.
26. Kalaiselvi VK, KR VC, Tirunagari S, Hariharan S, Krishnamoorthy M. Empowering smart traffic avoidance using Natural language processing. In 2025 IEEE International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE). 2025 Jan 16; 1–5.
27. Li Y, Ding H, Chen H. Data processing techniques for modern multimodal models. In 2024 IEEE Thirteenth International Conference on Image Processing Theory, Tools and Applications (IPTA). 2024 Oct 14; 1–6.
28. Glushkova TA, Valchev EV, Krasteva IK. Personalization of Lifelong Learning in School Educational Platform. In 2024 IEEE International Conference on Information Technologies (InfoTech). 2024 Sep 11; 1–4.
29. Murad DF, Toha M, Mayatopani H, Wijanarko BD, Heryadi Y, Dewi MA, Leandros R. Personalized recommendation system for online learning: An opportunity. In 2023 IEEE 8th International Conference on Business and Industrial Research (ICBIR). 2023 May 18; 128–132.
30. Xu H, Qin D, Liu C, Zhang Y. An improved dynamic model updating method for multistage gearbox based on surrogate model and sensitivity analysis. IEEE Access. 2021 Jan 21; 9: 18527–37.
31. Guo Z, Wang J, Tong Y, Zhang C, Liang B, Ma S. Technologies of distributed data stream processing based on big data. In 2021 IEEE International Conference on Computer Technology and Media Convergence Design (CTMCD). 2021 Apr 23; 244–247.

