Generative AI-based Approach to Concept Drift Generation in Streaming Text Data

BORIS BELOV, PETER PANFILOV
Department of Business Informatics,
HSE University,
Myasnitskaya St. 20, Moscow 101000,
RUSSIAN FEDERATION

Abstract: - Real-time analysis of text streams is crucial for industrial and business processes and scenarios. It is expected to be one of the important future research topics in the text processing and understanding domain. Analysis of text data is based on the use of pre-trained machine learning/data mining (ML/DM) models that may demonstrate performance degradation over time due to the drift in text data. The problem of tracking drift in data and quickly retraining a model in response to changes in the operational environment represents a great challenge in product model environments. We discuss and evaluate an approach to artificially generating concept drift aimed at providing test data for evaluating model performance and improving its accuracy. Existing methods for generating concept drift in text streams are limited to specific domains and are not universally applicable. This paper explores approaches for generating concept drift in text streams using the latest developments in generative artificial intelligence (GenAI) such as Large Language Models (LLMs). Two methods for generating concept drift with LLMs are proposed and compared to existing techniques. The comparison demonstrates that concept drift generation using LLMs is more effective than traditional methods. Additionally, LLMs can rapidly produce complex concept drift scenarios that are significantly more challenging to generate with standard approaches.

Key-Words: - generative AI, large language model, concept drift, data drift generation, drift detection methods, text data streams, lifelong machine learning, model retraining.

Received: March 17, 2024. Revised: August 16, 2024. Accepted: September 7, 2024. Available online: October 21, 2024.

1 Introduction

In the modern world, a huge amount of text data is produced. This data has enormous potential for solving various problems, and their processing can be very useful. For example, posts from social networks can be considered social sensor, [1]. Analyzing such data can help predict public opinion, understand user behavior, and simplify production processes.

Text data in many industries is streaming data. It arrives sequentially in small portions in a timely manner. A feature of streaming data is its sequence, the ability to sort by the time of arrival of data, and potential infinity, [2].

Pre-trained models are used to analyze such data, which makes it possible to speed up and qualitatively improve the process of extracting useful information. When working with streaming data, the model must be robust to the speed at which data appears, its volume, and potential changes in the data.

When training a model, it is assumed that the training set is a perfect reflection of the population of data. Due to time gaps or changes in environment and operational circumstances data may vary. In this case, we are facing the phenomenon of data drift that can negatively affect the performance of the pre-trained model and lead to incorrect interpretation of the data obtained. When working with text data, tracking drift and quickly retraining a model represent a great challenge in product model environments.

There is a large number of works devoted to the problem of detecting drift in text data based on the analysis of the decrease in the quality of the model, but at the same time, there are few works devoted to modeling drift and, accordingly, generating artificial drift. This applies both to the data drift itself and even more to the concept drift.

This work is designed to fill the gap in the field of synthetic data intended for the training and retraining of models used to process streaming text data.

The novelty of the presented research lies in the fact that Generative AI is supposed to be used to simulate drift in text data and generate drift.

2 Drift Problem Formulation

By drift in text data we understand the deviation of real data, that is, data at the model use stage, from the data that was used to test and train the model. The most general classification of drift in text data includes Data drift and Concept Drift.

2.1 Data Drift vs Concept Drift

Using some basics and terminology from probability theory we can formally define data drift as follows. Let $P(X)$ be the unconditional probability of obtaining a feature vector X in the input data set, and $P(y|X)$ is the conditional probability of obtaining values of the target variable y in the presence of input features X:

$$P(y|X) = \frac{Number\ of\ times\ X\ is\ obtained}{Total\ number\ of\ possible\ feature\ vectors} \quad (1)$$

$$P(y|X) = \frac{P(y \cap X)}{P(X)} \quad (2)$$

A data drift is a feature instance drift that involves a change in $P(X)$ caused by a change in the input data set. In this case, the dependence of the target variable y on the feature vector X remains unchanged. However, a change in probability $P(X)$ may entail a change in probability $P(y|X)$.

To identify data drift, it is first necessary to identify a change in the X-vector data distribution.

In turn, a concept drift is a change in $P(y|X)$, i.e. a change in the dependence of the target variable y on the feature vector X. Here it is assumed that the distribution of features itself may be unchanged, however, the patterns revealed by the machine learning model begin to emerge less reliably. As a result, the trained models become ineffective.

The fundamental relationship defining concept drift is characterized by the change in the distribution of X and y relative to the initial time t after a long period Δt:

$$P(y|X, t) \neq P(y|X, t + \Delta t) \quad (3)$$

A discovery of concept drift involves identifying changes in the distribution of the target variable y and/or changes in the dependence of y on X.

A concept drift can be exemplified by the Covid-19 pandemic situation, as it affects the economic situation, human interaction, and other important relationships.

2.1.1 Drift Detection

To determine the presence of drift in data special algorithmic solutions using statistical or other comparison methods can be used to relate newly received data with a historical one. Among those methods, the methods based on sequential analysis, model-based methods, and methods based on distribution estimation can be mentioned.

Research in the field of drift detection is carried out across a wide range of industries and disciplines, including, for example, medicine [3] and retail, streaming text data analysis (applications like Twitter) [4] recommendation systems [5], etc.

When working with text data, different algorithms can be used to detect drift. In particular, the Kolmogorov-Smirnov test is one of the basic detection methods, [6]. Another popular method is the Hoeffding Drift detection method (HDDM) [7], one of the groups of Window-based methods, according to the classification in a paper, [8]. The group of these methods is quite extensive and includes the Drift Detection Method [9], the Early Detection Method [10] and the Exponentially Weighted Moving Average [11], among others.

2.1.2 Drift Adaptation

A straightforward way to adapt the model to data drift is continuous learning. There are three main approaches to lifelong learning:

- *Retraining*. The ML algorithm is "completely" or "fully" retrained when a large volume of new data arrives, [12]. This approach significantly complicates the infrastructure requirements.
- *Incremental training*. The ML model is "immediately" or "permanently" retrained when new data arrives. It may introduce significant changes to the model weights due to a small increase in the amount of atypical data, [13].
- *Batch training*. When a large volume of new data arrives, the model is "partially" retrained. With this approach, an additional delay between training iterations is introduced.

2.1.3 Drift Generation

This is one of the under-researched areas, while a good solution to the problem of drift generation can help in a number of important modeling tasks. Among them, the most important are testing the model for resistance to drift at the development stage, developing new indicators to assess the quality of models, and planning time, computational, and financial resources for re-training of models.

Here we can mention two major classes of existing methods, namely: methods based on changing the data class [14] and methods based on changing words to opposite meanings, [15].

The former includes the class swap method, when classes change among themselves after a certain period of time, the class shift method which is similar to the previous one but with gradual change in classes, and the Time-slice Removal method when selected time periods are removed randomly from the original dataset. In a study [14], the methods discussed above were used on two different types of text datasets. Results demonstrated that all methods were found to be viable for creating a drift in data.

The latter has the only representative method known in the literature as an “adjective swap”. The general idea of the method is as follows: When working with text data, we can replace an adjective with its opposite meaning, which will distort the original meaning of the text and create a change in the text data.

3 Drift Problem Solution

The main hypothesis of our study, the preliminary results of which are presented in this paper, is as follows: Taking into account the development of a new type of Generative AI models that allows generating text using a pre-trained Large Language models (LLM), we assume that it is possible to use Generative AI-based solutions for faster, cheaper and higher-quality drift generation in text data.

3.1 GenAI Solutions for Text Data

The utilization of Generative AI solutions such as LLMs in scientific research is increasingly documented in scientific literature. A comprehensive study published in 2024 systematically categorizes the applications of LLMs for scientific purposes, [16]. Drawing from an analysis of over 400 scientific papers, the author identifies six primary research areas where LLMs are applied:

- Data generation and argumentation.
- Natural language processing and text analysis.
- Enhancement of machine learning models.
- Domain-specific applications.
- Security, privacy, and ethics.
- Emerging technologies and social impact.

Each category includes several sub-questions pertinent to its domain. Of particular interest in our research is the application of LLMs for data generation, encompassing two subcategories:

- synthetic data creation, which involves generating artificial data using LLMs in cases where real data is limited or subject to stringent data privacy regulations, and
- text data expansion and enhancement, which focuses on utilizing LLMs to generate new textual content and strengthen existing datasets to enhance machine learning models.

Research in this field explores diverse aspects of data generation, including efforts to synthesize text data. For instance, in one study, LLMs were employed to generate movie descriptions aimed at attracting viewers [17], demonstrating comparable quality to descriptions generated through traditional manual methods.

While researchers emphasize the capability of LLMs to address data imbalance issues and augment training data for machine learning models [18], there is limited discussion on their specific use for generating drift data and exploring the potential applications in this context.

However, it is crucial to consider that employing LLMs entails risks and necessitates thorough monitoring of the generated data. Recent studies, such as one conducted in 2023 by researchers from Berkeley and Stanford University [19], underscore changes in model behavior that warrant careful attention. Furthermore, in certain domains, generating data using GPT may introduce risks related to data accuracy, biases, and ethical considerations.

An analysis of the literature on the use of LLM for working with text data shows that with a huge bunch of work on the use of LLM in generating text data, there is a lack of work on handling drift in text data using LLM. In our study, we suggest that leveraging LLMs for generating data drift represents a promising avenue for future research.

3.2 GenAI Solutions for Russian Text Data

In our study of text data drift generation, several LLMs were considered candidates for integration into concept drift-generating solutions.

First of all, since the dataset chosen for the study is in Russian, it was necessary to select a model that performs well in the Russian language. Creating a custom LLM was not considered, as it would complicate the drift generation process. When selecting a pre-trained LLM, several parameters were taken into account, including operational speed, request limits, and the absence of hallucinations—meaning the text generated by the model remains relevant to the given topic. The most well-known large language models in Russian are Yandex GPT [20] and GigaChat, [21].

We made a comparison between these two models. In terms of technical characteristics (working speed and request limits), the two models demonstrated similar performance. However, during test queries on the selected data, it was found that GigaChat has a tendency to hallucinate. For example, when given the prompt ”replace adjectives with antonyms” in a text related to a cafe, the model could produce irrelevant content, such as the rules of tennis. At the same time, Yandex GPT did not exhibit such hallucinations.

3.3 GenAI-based Concept Drift Generation

In this work, we propose two new methods for generating concept drift in data:

- Adjective swap with LLM, which is replacing words with opposite ones using question-answering systems like GPT.
- LLM-assisted text shortener. Sentence transformation that changes source texts using prompts in question-answering systems in any given direction (for example, reducing the length of a sentence while maintaining the meaning, etc.).

As a proof of concept, we designed and conducted experiments with the Yandex GPT Lite model by comparing previously known approaches to generating data drift (e.g., class shift and adjective swap) with the proposed approach to generating data drift using the LLM-based solution.

The Yandex GPT Lite LLM was the focus of our experimental study. It offers a cloud API that supports prompt queries. The execution results can be obtained either synchronously or asynchronously.

In synchronous mode, the LLM returns results in response to the request immediately, with a throughput capacity of 500 requests per hour.

In asynchronous mode, a request ID is returned in response to the prompt query, which can then be used to retrieve the execution result via a separate API. The execution result appears within an hour after the request is processed. Asynchronous requests have a higher limit of 20,000 requests per hour, and the cost per token in asynchronous mode is half that of synchronous mode.

In both modes, users are limited to sending one request at a time, a restriction applicable to non-corporate Yandex GPT users.

Synchronous mode was employed for testing prompt requests, while asynchronous mode was utilized for data conversion.

4 Experiments and Results

In the situation of insufficient publications on the use of LLM to generate a concept drift, we conducted experiments and compared the results of our solutions based on LLM with existing ones such as class shift and adjective swap. Queries for which the LLM showed the best results were manually selected.

4.1 Experimental Design and Setup

Workflow diagram for conducting experiments is presented in Figure 1. It includes the four major steps:

- Add drift to test data.
- Apply concept drift detection method to test data.
- Test the model on 70% of test data with concept drift.
- Test the model on the remaining 30% of test data with concept drift.

Experimental setup (HW/SW testbed) used in experiments included:

- M1 Pro laptop with macOS Monterey (16GB Ram, 1TB storage).
- Python 3.7. with packages: pandas (0.23.4), matplotlib (3.5), numpy (1.16.0), PyMultiDictionary (1.2.4).

The overall design of the experiment included a comparison of four approaches to generating concept drift in the text, namely:

- *Class shift*. During initial data processing, ratings 1-4 were classified as negative, and 5 as positive. For drift generation, a score of 4 will be reclassified as positive.
- *Adjective swap*. Adjectives and adverbs in the text will be replaced with their antonyms.
- *Adjective swap with LLM*. Words in the text will be replaced with their antonyms using an LLM.
- *LLM-assisted text shortener*. Negative reviews tend to be longer than positive ones, potentially making the model sensitive to the length of negative reviews. To generate a concept drift, long texts will be shortened to 240 characters without losing the original meaning.

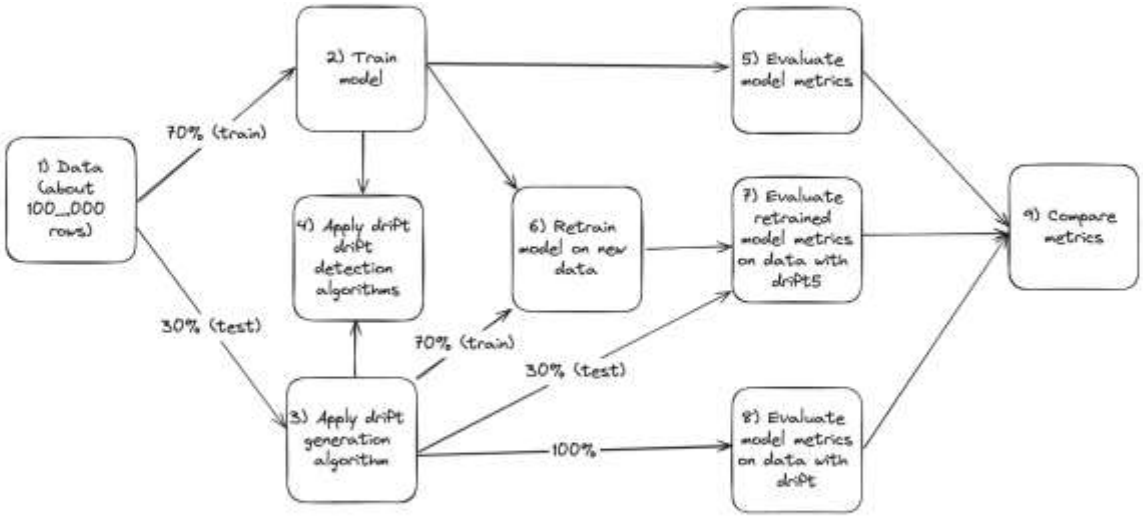

Fig. 1: Experimental design

4.2 Data for Experiments

To conduct the experiment, we need a text dataset with marked features. It was decided to conduct an experiment with Russian language data. The following criteria were used to choose a dataset:

- *Number of rows*. The more rows we have in the dataset, the less likely it is that the model performs well on the training data.
- *Average length of texts*. If the texts are very short, then even a small change can significantly change their meaning.
- *Data relevance*. The data must have value for use in solving business problems or research purposes.

4.2.1 Experimental Dataset

Experiments with LLM-based data drift generation were conducted using the Geo Reviews Dataset with reviews of organizations from Yandex Maps for August 2023. The overall text volume of the dataset is 500.000 reviews in the Russian language with ratings from 1 to 5. The dataset contains reviews for many types of organizations - restaurants, medical institutions, beauty salons, etc. To make the data more uniform, 100.000 reviews for cafes and restaurants were selected from the dataset (Figure 2).

The dataset contains the following columns:

- Organization address.
- Organization name.
- Organization rating (from 1 to 5).
- Type of activity of the organization.
- User review of the organization.

The task was reduced to a classification problem with two classes of positive and negative reviews: ratings 1-4 are considered "positive", and rating "5" is negative. As a result, the two classes have a ratio of 25 to 75. Let's look at the data in the column with the text of the review as a whole and broken down into "positive" and "negative" classes (Table 1).

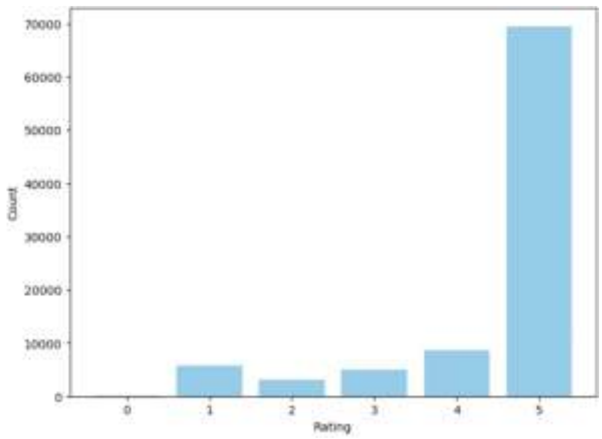

Fig. 2: Distribution of reviews by ratings

Table 1. Comment Length Statistics

Metric	Review length		
	All data	***Positive reviews***	***Negative reviews***
Average	282.5	235	427
25th percentile	89	84	120
50th percentile	206	187	323
75th percentile	554	432	855

Figure 3 shows that the positive reviews are on average almost twice as long as negative ones.

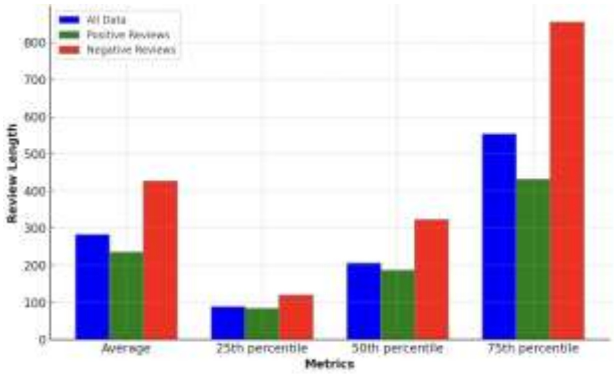

Fig. 3: Comparison of the reviews by length

4.2.2 Data Preparation

To conduct the experiment, the dataset was divided into parts in a ratio of 70 to 30. A larger part will be used to train the model. A smaller part will be used to generate data drift and check for changes in the quality of model predictions.

Text data was specially prepared before training:

- Punctuation and non-letter characters removed from the text.
- Stop words removed from the text.
- Lemmatization of words applied.
- The texts vectorized using TfidfVectorizer.

The presence of concept drift in the data is detected using the Kolmogorov-Smirnov test and the HDDM method. The initial data contains no drift, so the resulting values will serve as a baseline for subsequent evaluations.

The baseline for the Kolmogorov-Smirnov test is 0.165. To confirm the presence of drift, the value must be less than 0.05.

The baseline for the HDDM test is 0.403. To confirm the presence of drift, the value must be greater than the baseline.

Additionally, when testing hypotheses, we will consider the F1 score.

The Multinomial Naive Bayes algorithm, [22] was used as the machine learning model. This algorithm supports incremental training, performs well with numerical vectors containing independent features, and handles sparse data effectively.

Table 2 presents the baseline results obtained after training model on test data without concept drift.

Table 2. Baseline Model Metrics

Class	*Precision*	*Recall*	*F1 score*	*Accuracy*
Negative	0.90	0.58	0,71	0,89
Positive	0.88	0.98	0,93	

4.3 Experiments

Four experiments were conducted to collect data on four techniques to generate data drift in the text, namely:

- Class shift.
- Adjective swap.
- Adjective swap with LLM.
- LLM-assisted text shortener.

4.3.1 Class Shift

To perform class shift reviews with a rating 4 were reclassified as “positive” in test data.

The Kolmogorov-Smirnov test and the HDDM method do not make sense for this drift generation method, since the original text data has not changed.

The use of class shift led to an unexpected result – the accuracy of predictions increased. This result can be explained by a feature of the data – reviews with a rating of 4 are very similar to reviews with a rating of 5 and there are many of them, so before the class shift the model often made the mistake of taking reviews with a rating of 4 as positive. After the class shift, such errors became correct predictions, which had a positive effect on the quality of model predictions.

4.3.2 Adjective Swap

Adjectives and adverbs accounted for approximately 15% of the total number of unique words in the dictionary. Due to some words containing typos, after lemmatization, the same word with and without a typo can be considered two different words. To fully implement this approach, it is necessary to correct typos in the comments.

To implement this drift method, all adjectives and adverbs were collected from the comments, resulting in 5,224 adjectives and 1,443 adverbs. The PyMultiDictionary library was used to search for antonyms.

None of the available Python libraries support searching for antonyms in Russian, so a translation-based approach was employed in our solution as follows:

- Translate the word into English (among the possible translations, the most popular one is selected).
- Find the antonym of the word in English (among the possible antonyms, the most popular one is selected).
- Translate the antonym of the word back into Russian (among the possible translations, the most popular one is selected).

The Russian language is rich in words that are similar in meaning but used in different contexts, so searching for antonyms without considering the context can lead to meaningless phrases. For example, for the word ”fantastically,” the possible antonym could be ”frequently,” although in the phrase ”fantastically delicious,” this substitution would be inappropriate.

The Kolmogorov-Smirnov test in our experiments with adjective swap showed a result of 0.2, and the HDDM method showed a result of 1.08. According to the Kolmogorov-Smirnov test, there is no drift, but HDDM has increased from the base value, which can be considered a sign of drift. F1 score worsened for negative reviews.

The use of Adjective swap generated a concept drift, which significantly affected the quality of predictions of negative reviews and had a minor effect on the quality of predictions of positive reviews. After additional training on the model, the quality of predictions of negative reviews almost reached the baseline model. This heterogeneous behavior can be explained by:

- A small share of adjectives in the dictionary.
- Ignoring the context of the use of a word when replacing it with an antonym.
- Presence of typos in the text.
- Testing the model on test data.
- Testing the model on the remaining 30% of test data.

4.3.3 LLM-assisted Adjective Swap

A new approach to replacing words with antonyms using LLM was introduced and studied in our research. At the core of this approach, the YandexGPT Lite large language model is applied to convert comments when using it with the prompt ”Rewrite the text by replacing words with antonyms, maintaining the length of the text.” This prompt was selected on the basis of the results of experimentations.

The Kolmogorov-Smirnov test showed a result of 0, and the HDDM method showed 1.24, which indicates the presence of Concept drift in text data. F1 score dropped significantly for both positive and negative reviews. After training, F1 scores approached the baseline values.

4.3.4 LLM-assisted Text Shortener

As it was demonstrated earlier, upon examining the data, it was discovered that negative reviews are, on average, almost twice as long as positive ones. To equalize the length of positive and negative reviews, a large language model was used. The prompt ”Shorten the text to 200 characters while preserving the meaning” was selected for this purpose. This command was applied only to reviews longer than 240 characters.

As a result, the length of the converted reviews varied from 150 to 240 characters and, on average, was significantly shorter than the original.

The Kolmogorov-Smirnov test showed a result of 0, and the HDDM method showed a result of 0.57, which indicates the presence of Concept drift. The F1 score deteriorated for negative reviews, as they were mostly the only ones that changed. After training the F1 score was restored to baseline. Together, all this demonstrates the presence of Concept Drift.

4.4 Summary of Experimental Results

As a result of the experiments with 4 methods of generating concept drift, the following conclusions can be drawn:

- Class shift is not always successfully used to create concept drift. In the experiment, we initially classified only reviews with a rating of 5 as positive. When using class shift, rating 4 also began to be classified as positive, which led to an improvement in the quality of the model. The class shift can lead to improved model quality if we shift a class that was classified incorrectly before the shift.
- Adjective swap did not create data drift for all classes. Replacing words with antonyms does not work accurately, since words can be located in the context of other words and contain errors. In the case of the Russian language, this task becomes more complicated, since there are no tools for the Python language that support searching for antonyms for the Russian language.
- Replacing words with antonyms using LLM proved to be significantly better than traditional Adjective swap. The explanation is that LLM takes into account the context of words and is not sensitive to errors in the text.
- Reducing the number of characters in long reviews to average values in the dataset using LLM has been shown to be effective in creating data drift in the data.

Table 3 summarizes the results of experiments with different variants of the data drift generation.

Table 3. Comparison of Four Approaches

Metric/Method	Class shift	*Adjective swap*	*LLM Adjective swap*	*LLM Shortener*	*Baseline*
Positive F1 score	0,71	0.48	0.38	0,62	0,71
Negative F1 score	0,96	0.90	0.45	0,92	0,93
Accuracy	0,93	0,82	0,41	0,86	0,89
Positive F1 score retraining	0,78	0,67	0,62	0,71	
Negative F1 score retraining	0,96	0,92	0,91	0,93	
Accuracy retraining	0,94	0,88	0,85	0,89	

5 Discussion

The generative AI concept refers to different forms of content generation, including text data where large language models or LLMs as a special kind of generative AI application play an important role of foundation models – a basis for text processing tasks of various kinds. We propose to use LLM in solving drift problems in text data which is instrumental in the performance degradation of machine learning models that deal with streaming text data.

Our experiments with different data and concept drift generation methods in comparison with our own LLM-based methods demonstrated several advantages of LLM-based drift generation over traditional methods:

- There is no need to use complex natural language processing tools. For example, for the Russian language, it was not possible to find a Python library for searching for antonyms of words.
- There is no need to correct typos and errors in the text since LLM is not sensitive to them.
- You can test hypotheses that would be very difficult to implement using standard methods, such as shortening the text.
- This approach is low-impact in terms of time and cost. Processing one thousand reviews cost about 30-50 rubles (~50 cents) and took several minutes. When using a corporate payment plan (for cloud services, such as HPC-as-a-service), processing can be performed in parallel in multiple parallel threads, which will speed up the process.

Overall, the comparison of 4 drift generation methods demonstrates that concept drift generation using LLMs is more effective than traditional ones. The LLM-based approach to adjective swap demonstrates advantages over the traditional adjective swap in that LLM is capable of selecting antonyms in the context of the word's use and is not sensitive to typos in the text. The experiments with LLM-based text shortener show that using LLM it is possible to successfully simulate complex concept drift scenarios, the modeling of which would be extremely difficult using standard approaches.

However, the generalizability of the results is limited by the size of the dataset and the use of only one machine learning model in experiments. To obtain more convincing results, it would be necessary to expand the experimental plan to include larger datasets to retrain the model based on drift-affected data, experimenting with various machine learning algorithms, and testing LLM-enabled methods on multilingual datasets.

Since we aim to improve the quality of models, which deteriorates over time, it would also be useful to examine whether the proposed drift generation methods improve the performance of machine learning models during retraining and evaluate the effectiveness of expanding training data sets using LLM-generated text, and if so, how.

6 Conclusion

In this paper, we explored the use of LLM to generate concept drift in textual data for use in retraining models in order to improve model quality and/or mitigate the effects of performance degradation over time due to data drift.

Our experiments with various "traditional" methods of generating data drift and concept drift in text data in comparison with our own LLM-based methods have demonstrated a number of advantages of LLM-based drift generation over existing ones. Firstly, the experimental results showed that the proposed approaches using LLM successfully generate drift in text data. Secondly, these methods make it possible to implement various drift scenarios with less time and resources. In addition, LLM can quickly create complex concept drift scenarios that are much more difficult to create using standard approaches.

Unlike traditional methods described in the literature, which have limited applicability outside of research laboratories, the LLM-based methods presented in this paper open up new possibilities for generating drifts in text data and can offer practical solutions to solve various business problems. One important application is the evaluation of models designed to recognize toxic comments. The presented methods allow for assessing how variables such as comment length and slang usage affect the performance and quality characteristics of the model.

In the field of education, these methods are of considerable value due to the lack of textual datasets demonstrating drift in open-source repositories. This disadvantage creates problems for studying drift and understanding the behavior of the model in response to drift. The proposed methods may facilitate the generation of drift in text datasets, thereby improving the study of drifts and the performance of models in various text datasets.

Declaration of Generative AI and AI-assisted Technologies in the Writing Process

During the preparation of this work the authors used Yandex Translate a machine translation system developed by Yandex in order to improve the readability and language of their paper. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

References:

[1] A. Suprem and C. Pu, "ASSED: a Framework for Identifying Physical Events Through Adaptive Social Sensor Data Filtering," in *Proc. 13th ACM Int. Conf. Distributed and Event-based Systems*, Darmstadt Germany, June 24 - 28, 2019, pp. 115–126. doi: 10.1145/3328905.3329510.

[2] A. Bifet, R. Gavalda, G. Holmes, and B. Pfahringer, *Machine Learning for Data Streams: with Practical Examples in MOA.* MIT Press, 2018. doi: 10.7551/mitpress/10654.001.0001.

[3] K. Rahmani, R. Thapa, P. Tsou, S. C. Chetti, G. Barners, C. Lam, and C. F. Tso, "Assessing the effects of data drift on the performance of machine learning models used in clinical sepsis prediction," *Int. J. Med. Info.*, Volume 173, May 2023, 104930. doi: 10.1016/j.ijmedinf.2022.104930.

[4] A. Bechini, A. Bondielli, P. Ducange, F. Marcelloni, and A. Renda, "Addressing Event-Driven Concept Drift in Twitter Stream: A Stance Detection Application," *IEEE Access*, vol. 9, pp. 77758-77770, 2021. doi: 10.1109/ACCESS.2021.3083578.

[5] A. D. Vinisky, J. P. Barddal, A. de S. Britto Jr., F. Enembreck, and H. V. A. Campos, "A case study of batch and incremental recommender systems in supermarket data under concept drifts and cold start," *Expert Syst. Appl.*, Volume 176, February 2021, 114890. doi: 10.1016/j.eswa.2021.114890.

[6] R. Feldhans, A. Wilke, S. Heindorf, and M. H. Shaker, "Drift Detection in Text Data with Document Embeddings," in Proc. 22nd Int'l Conf. *Intelligent Data Engineering and Automated Learning – IDEAL* 2021, Manchester, UK, November 25–27, 2021, pp. 107-118. doi: 10.1007/978-3-030-91608-4_11.

[7] I. Frias-Blanco, J. del Campo-Avila, G. Ramos-Jimenez, R. Morales-Bueno, A. Ortiz-Diaz, and Y. Caballero-Mota, "Online and non-parametric drift detection methods based on Hoeffding's bounds," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 3, pp. 810–823, Mar. 2015. doi: 10.1109/TKDE.2014.2345382.

[8] J. Gama, I. Zliobaite, A. Bifet, M. Pechenizkiy, and A. Bouchachia, "A survey on concept drift adaptation," *ACM Comput. Surv.*, vol. 46, no. 4, Article 44, pp.1-37, 2014. doi: 10.1145/2523813.

[9] J. Gama, P. Medas, G. Castillo, and P. Rodrigues, "Learning with drift detection," in Proc. *17th Brasilian Symposium on AI - SBIA 2004*, Sao Luis, Maranhao, Brasil, September 29-October 1, 2004. A. L. C. Bazzan and S. Labidi, Eds. Heidelberg: Springer, 2004, pp. 286–295. doi: 10.1007/978-3-540-28645-5_29.

[10] M. Baena-Garcia, J. del Campo-Avila, R. Fidalgo, A. Bifet, R. Gavalda, and R. Morales-Bueno, "Early drift detection method," in Proc. *4th Int. Workshop Knowl. Discovery from Data Streams*, Berlin, Germany, September 18, 2006, pp. 77–86.

[11] G. J. Ross, N. M. Adams, D. K. Tasoulis, and D. J. Hand, "Exponentially weighted moving average charts for detecting concept drift," *Pattern Recogn. Lett.*, vol. 33, no. 2, pp. 191–198, Feb. 2012. doi: 10.1016/j.patrec.2011.08.019.

[12] N. Sanjay, "Continuous Training of ML models," *Medium.*, [Online]. https://medium.com/@nagasanjayvijayan/continuous-training-of-ml-models-7d8acaf44dda (Accessed Date: September 27, 2024).

[13] A. Bifet and R. Gavalda, "Adaptive learning from evolving data streams," in Proc. *8th Int. Symp. Intell. Data Analysis*, IDA 2009, Lyon, France, August 31 – September 2, 2009, pp. 249-260. doi: 10.1007/978-3-642-03915-7_22.

[14] C. M. Garcia, A. L. Koerich, A. S. Britto Jr., and J. P. Barddal, "Methods for Generating

Drift in Text Streams," *arXiv* preprint, [Online]. https://arxiv.org/html/2403.12328v1#bib.bib12 (Accessed Date: September 27, 2024).
[15] F. Bravo-Marquez, A. Khanchandani, and B. Pfahringer, "Incremental Word-Vectors for Time-evolving Sentiment Lexicon Induction," *Cogn. Comput.*, vol. 14, no. 5, 2022, pp. 425–441. doi: 10.1007/s12559-021-09831-y
[16] F. Sufi, "Generative pre-trained transformer (GPT) in research: a systematic review on data augmentation," *Information*, 15(2), 99; MDPI, 2024.doi: https://doi.org/10.3390/info15020099.
[17] A. Acharya, B. Singh, and N. Onoe, "LLM Based Generation of Item-Description for Recommendation System," in Proc. *17th ACM Conf. Recommender Systems*, Singapore, September 18-23, 2023, pp. 1204–1207. doi: 10.1145/3604915.3610647.
[18] C. Suhaeni and H. S. Yong, "Mitigating Class Imbalance in Sentiment Analysis through GPT-3-Generated Synthetic Sentences," *Appl. Sci. 2023*, vol. 13, no. 17: 9766. MDPI, 2023. doi: 10.3390/app13179766.
[19] L. Chen, M. Zaharia, and J. Zou, "How is ChatGPT's behavior changing over time?" *arXiv* preprint, 2023, [Online]. https://arxiv.org/pdf/2307.09009v2 (Accessed Date: September 27, 2024).
[20] What is YandexGPT, 2023, [Online]. https://ya.ru/ai/gpt-2 (Accessed Date: September 27, 2024).
[21] Gigachat, 2023, [Online]. https://giga.chat/ (Accessed Date: September 27, 2024).
[22] MultinomialNB, sklearn.naive bayes, [Online]. https://scikit-learn.org/stable/modules/generated/sklearn.naive_bayes.MultinomialNB.html (Accessed Date: September 27, 2024).

Contribution of individual authors to the creation of a scientific article (ghostwriting policy):

- Boris Belov implemented software tools and applications designed and developed for experimentation with data drift generation methods, carried out programming and experimenting with software apps, and collected experimental data.
- Peter Panfilov carried out the overall coordination of the experimental work including the design of the experiments and methodology for experimental data collection and processing, and was responsible for the analysis and interpretation of the results of the experimental work.

Sources of funding for research presented in a scientific article or scientific article itself:
No funding was received for conducting this study.

Conflict of interest:
The authors have no conflicts of interest to declare.