Handling Large Contexts in Large Language Models: Strategies Beyond Model Limits

Lucas Nóbrega and Luísa Marschhausen

Programa de Engenharia de Sistemas e Computação
Universidade Federal do Rio de Janeiro (UFRJ)
Rio de Janeiro – RJ – Brazil

e-mail: {lucasnobrega,lmarschh}@cos.ufrj.br

September 2023

Abstract

Within the domain of Large Language Models (LLMs) like GPT, and LLaMA, context size limitations pose significant challenges, especially when the text corpus surpasses the model's maximum input size. This paper presents a review, subsequently introducing heuristics, techniques, and algorithms tailored for processing large contexts. Drawing inspiration from parallel models and functional programming operations, we propose a bifurcated solution: text partitioning and its subsequent utilization.

1 Introduction

The most significant advance in natural language processing (NLP) is the emergence of Large Language Models (LLMs) (Brown et al., 2020; Devlin et al., 2018; Conneau et al., 2019; OpenAI, 2023). LLMs, exemplified by models like BERT (Devlin et al., 2018), different versions of GPT (Radford, Narasimhan, et al., 2018a; Radford, Wu, et al., 2019; Brown et al., 2020; OpenAI, 2023), and LLaMA (Petroni et al., 2019), have established new benchmarks in a multitude of NLP tasks, from machine translation to sentiment analysis (Vaswani et al., 2017; Yang et al., 2019). GPT-4 achieved human-level proficiency in more than 30 professional and academic exams (OpenAI, 2023). Their vast architectures, which often encompass billions of parameters, underscore their ability to discern and model intricate linguistic nuances (Radford, Narasimhan, et al., 2018b).

However, inherent challenges accompany these models. A salient challenge is the limitation in context size (Liu et al., 2019). Here, "context" denotes the segment of text that a model can ingest and process in a single iteration. Due to both architectural and computational constraints, this context is bound by an upper limit. Consequently, when confronted with voluminous texts or corpuses that exceed this threshold, the utility of LLMs becomes constrained (Beltagy et al., 2020).

This limitation is palpable in scenarios that frequently encounter extensive texts, such as academic research, conversational analysis, or exhaustive summarization tasks (Zhang et al., 2019). Whether it is a researcher analyzing a scientific article, an analyst parsing lengthy conversation logs, or a content creator attempting to summarize a lengthy piece, the sheer volume of text can easily outstrip the context size of contemporary LLMs.

Compounding this challenge is the diversity inherent in texts and corpuses. Texts can vary in size, from those slightly exceeding the LLM's limit to those so expansive that they necessitate intricate partitioning or sampling techniques (Küppers et al., 2018). Moreover, the variety of tasks these texts can undergo further magnifies the challenge. From monolithic corpus analysis to granular, text-specific operations, the solutions differ markedly (Zhao et al., 2019).

The dual objective of this paper is to first provide an exhaustive exposition of the context-size challenge and subsequently to introduce a gamut of techniques, heuristics, and algorithms tailored to navigate this challenge effectively. Drawing from parallel processing paradigms like map-reduce (Dean et al., 2008) and foundational tenets of functional programming (Bird et al., 1988), we posit a two-pronged approach: partitioning the text followed by strategic partition processing.

Subsequent sections will delve into our proposed methodologies, explore their applicability across diverse tasks, and evaluate their efficacy via rigorous experimentation. This exploration aims to empower researchers and practitioners with the requisite knowledge and tools to maximize the potential of LLMs, even when processing extensive texts.

2 Problem Formulation

Large Language Models, such as GPT and LAMMA, have demonstrated exceptional proficiency in understanding and generating text within a predefined context window. However, this inherent limitation presents a significant challenge when dealing with textual data that exceeds the model's context size. In this section, we formally define the problem and explore its various aspects, including scenarios where the text corpus significantly surpasses the LLM's context and cases where the mismatch is more modest.

2.1 Definition

Let us define the problem of extending the contextual understanding of LLMs formally. Consider a text corpus C consisting of a sequence of words $C = (w_1, w_2, \ldots, w_N)$, where N is the total number of words in the corpus. An LLM with a predefined context window size K is denoted as M.

The contextual understanding problem can be stated as follows: Given a text corpus C and an LLM M with a context window size K, we aim to create a mechanism that allows M to effectively process and comprehend contexts larger than K. This entails developing strategies and techniques to provide M with access to extended portions of C while maintaining its ability to generate coherent and contextually relevant responses.

2.2 Scenarios

To gain a deeper understanding of the problem, we consider two primary scenarios:

2.2.1 Scenario 1: Text Corpus Significantly Exceeds Context Size

In this scenario, the text corpus C is substantially larger than the LLM's context window size K. For example, C could be a lengthy research paper or a lengthy conversation transcript. In such cases, it becomes challenging for M to maintain context continuity and relevance.

2.2.2 Scenario 2: Text Corpus Modestly Exceeds Context Size

In contrast, in Scenario 2, the text corpus C is only slightly larger than the LLM's context size K. While M may be able to capture most of the context, there might be crucial information that lies beyond its reach. This scenario is common in tasks like long-form content generation or document summarization.

2.3 Importance

The significance of addressing the problem of extending contextual understanding in LLMs cannot be overstated. As textual data proliferates across diverse domains, from legal documents and academic literature to social media and customer reviews, LLMs must adapt to handle increasingly extensive contexts. This adaptation is vital for tasks such as document summarization, sentiment analysis of lengthy texts, and coherent responses in extended conversations.

In the following sections, we delve into various strategies, techniques, and algorithms to empower LLMs to overcome their context limitations and analyze textual data more comprehensively. We also explore the concept of sampling methods, which can be invaluable when dealing with massive datasets, and present mathematical formulations, algorithms and figures to illustrate our proposed solutions.

3 Sampling Strategies

3.1 Definition and Importance

The task of sampling in the context of LLMs and large textual datasets involves the careful selection of a subset of data points from a larger population. When dealing with vast collections of texts, direct processing might not be feasible due to computational limitations. Sampling becomes an invaluable tool in such scenarios. However, the choice of sampling method can influence the representativeness of the sample and, consequently, the accuracy of subsequent analyses.

3.2 Mathematical Formulation

Let S represent a set of data points or elements, where $S = (s_1, s_2, \ldots, s_n)$ and n is the total number of elements in the dataset. Our objective is to select a subset S_s from S such that S_s is a representative, reduced-size representation of the entire dataset.

Mathematically, the problem can be stated as:

$$S_s = \{s_i \mid i \in I, 1 \leq i \leq n_s\}$$

Where S_s is the sample set consisting of n_s sampled elements and I represents the index set of selected elements.

3.3 Strategies for Sampling

3.3.1 Uniform Sampling

Uniform Sampling, as represented in Figure 1, involves selecting text samples randomly, with each text having an equal probability of being chosen. This method is straightforward and ensures that the sample is unbiased in terms of the original distribution of the texts.

Figure 1: Example of Uniform Sampling ($n = 5$). Adapted from: `https://www.scribbr.com/methodology/sampling-methods/`

Algorithm 1: Uniform Sampling

Data: Collection C, Sample size n
Result: Sample S
begin
  $S \leftarrow$ Randomly select n texts from C;
  return S;

3.3.2 Stratified Sampling

Stratified Sampling, as represented in Figure 2, aims to ensure that certain subgroups of texts are adequately represented in the sample. This becomes crucial when there are characteristics that divide the text, like polarity in sentiment analysis, where naive random sampling might overlook minority classes.

3.3.2.1 Stratified Sampling with Known Class Labels

In scenarios where class labels are known, texts can be grouped by their label, and samples can be drawn proportionally from each group.

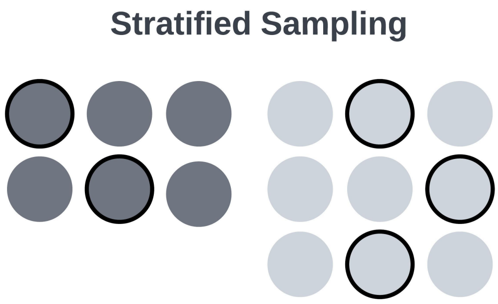

Figure 2: Example of Stratified Sampling ($n = 5$). Adapted from: https://www.scribbr.com/methodology/sampling-methods/

Algorithm 2: Stratified Sampling with Known Class Labels

Data: Collection C, Sample size n, Class labels L
Result: Sample S
begin
  Initialize empty set S;
  for *each label* l *in* L **do**
    $C_l \leftarrow$ All texts in C with label l;
    $n_l \leftarrow$ Proportional number of texts to sample from C_l;
    $S_l \leftarrow$ Randomly select n_l texts from C_l;
    Add S_l to S;
  return S;

3.3.2.2 Stratified Sampling with Computed Class Labels

When class labels are not provided, they must first be computed. After determining the class of each text, a sample can be drawn based on a predefined distribution.

3.4 Considerations for Sampling

When sampling from a large dataset, several considerations should be taken into account to ensure the representative nature and quality of the sample:

- **Randomness**: Random sampling helps avoid bias and ensures that each element in the dataset has an equal chance of being included. It's crucial to use proper randomization techniques.

Algorithm 3: Stratified Sampling with Computed Class Labels

Data: Collection C, Sample size n, Classification function F, Desired distribution D

Result: Sample S

begin

  Initialize empty set S and class-label map M;

  for *each text* t *in* C **do**

    $l \leftarrow F(t)$; `// Classify text`

    Add t to group M_l;

  for *each label* l *in* M **do**

    $n_l \leftarrow$ Number of texts to sample from M_l based on D;

    $S_l \leftarrow$ Randomly select n_l texts from M_l;

    Add S_l to S;

  return S;

- **Sample Size**: Determining the appropriate sample size is critical. A sample that's too small may not accurately represent the population, while an excessively large sample can be computationally expensive.
- **Stratification**: In some cases, it may be beneficial to stratify the sample, ensuring that it represents different subgroups within the dataset. This is particularly useful when dealing with imbalanced datasets or when specific subpopulations need special attention.
- **Temporal Considerations**: When working with time-series data, consider the temporal aspects of sampling. Random sampling over time may not capture trends or seasonality, so temporal stratification or other techniques might be necessary.
- **Bias and Unwanted Patterns**: Be aware of potential biases or unwanted patterns in the data that could affect the sampling process. Addressing such issues may require specialized sampling techniques.

4 Partitioning Strategies

4.1 Definition and Importance

The task of text partitioning arises from the need to split lengthy textual content into smaller segments that can be processed by LLMs without losing essential information. In the context of LLMs like GPT and LLaMA, which have a fixed maximum context size, partitioning becomes crucial when dealing with texts that exceed this limit.

However, the challenge is not just to split the text arbitrarily. Effective partitioning should ensure that the resultant segments are coherent and do not break the inherent structure or meaning of the text. This is particularly important for tasks that rely on contextual information, such as sentiment analysis or topic modeling.

4.2 Mathematical Formulation

Let T represent a text as a sequence $(t_1, t_2, \ldots, t_n)$ of tokens that needs to be partitioned. If m represents the maximum context size of an LLM, our objective is to partition T into smaller subtexts $T_1, T_2, \ldots, T_k$ such that each T_i is a sequence $(t_{i_1}, t_{i_2}, \ldots, t_{i_{|T_i|}})$ and contains at most m tokens, and k is minimized. The order of tokens in the original text T must be preserved in the partitioned subtexts.

Mathematically, the problem can be stated as:

$$\begin{aligned}
&\text{Minimize} \quad k \\
&\text{subject to:} \\
&\sum_{i=1}^{k} |T_i| = n \\
&|T_i| \leq m \quad \forall i \in \{1, 2, \ldots, k\} \\
&T = \bigcup_{i=1}^{k} T_i \\
&\text{If } t_{i_j} \text{ and } t_{i_{j+1}} \text{ are consecutive tokens in } T_i, \text{ then } i_j < i_{j+1}
\end{aligned}$$

Where $|T_i|$ represents the number of tokens in subtext T_i and constraints ensure the preservation of the order of tokens in the partitioned sequences consistent with their order in the original sequence T.

4.3 Strategies for Text Partitioning

4.3.1 Simple Splitting

In Simple Splitting, represented in Figure 3, the text is split every m tokens, without any regard for the structure of the content. While this method is computationally simple, it might break sentences or paragraphs, leading to a possible loss of meaning. Since we are doing text splitting because the text is too big to fit in the context, the effect should be marginal.

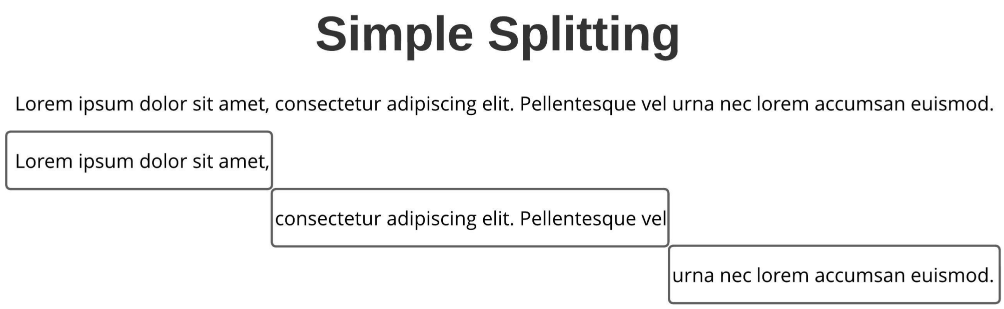

Figure 3: Example of Simple Splitting. ($P = 5$)

Algorithm 4: Equal Size Partitioning

Data: Text T, Partition size P

Result: Partitions of text T

partitions $\leftarrow$ empty list;
for $i = 0$ *to* $len(T)$ *by* P **do**
  partitions.append(T[i:i+P]);
return partitions;

4.3.2 Sliding Window Partitioning

In the sliding window approach, represented in Figure 4, the text is divided into overlapping segments, each of a fixed size w. The starting point of each segment is shifted by a fixed step size s, which determines the overlap between consecutive segments. This method ensures that context around segment boundaries is preserved to some extent, mitigating potential issues arising from abrupt splits in simpler methods.

4.3.3 Boundary-Based Partitioning

Boundary-Based Partitioning, as represented in Figure 5, involves splitting the text at natural boundaries, such as sentences or paragraphs, ensuring that the segments are coherent. This might lead to some segments being significantly smaller than m tokens, but the advantage is that the context remains intact.

Algorithm 5: Sliding Window Partitioning

Data: Text T, Window size w, Step size s

Result: Partitions of text T using sliding windows

```
partitions ← empty list;
for i = 0 to len(T) − w + 1 by s do
    partitions.append(T[i:i+w]);
return partitions;
```

Algorithm 6: Boundary-based Partitioning

Data: Text T as sequence of tokens $(t_1, t_2, \ldots, t_n)$, Maximum partition size m, Boundary function $B(t)$

Result: List of partitions P

```
begin
    Initialize an empty list P;
    Initialize an empty partition p;
    for each token t_i in T do
        Append t_i to p;
        if Boundary function B(t_i) is true OR |p| = m then
            Append p to P;
            Reset p to an empty partition;
    if p is not empty then
        Append p to P;
    return P;
```

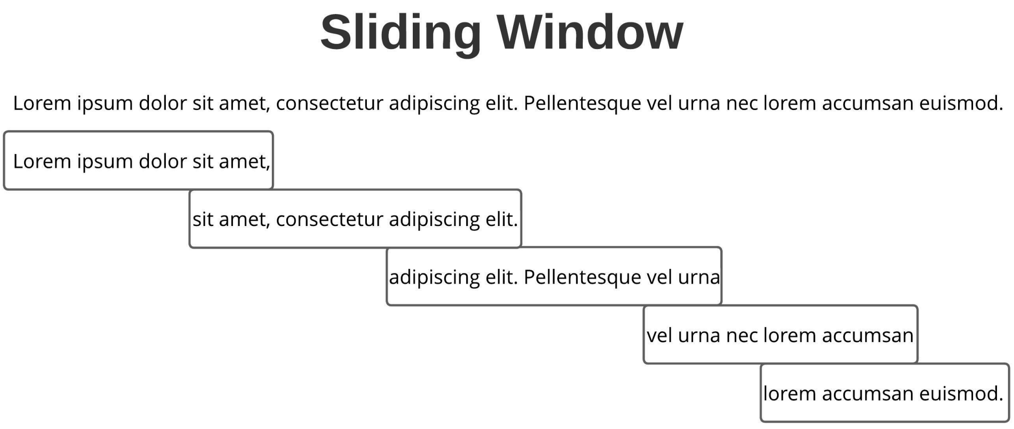

Figure 4: Example of Sliding Window Partitioning. ($w = 5, s = 3$)

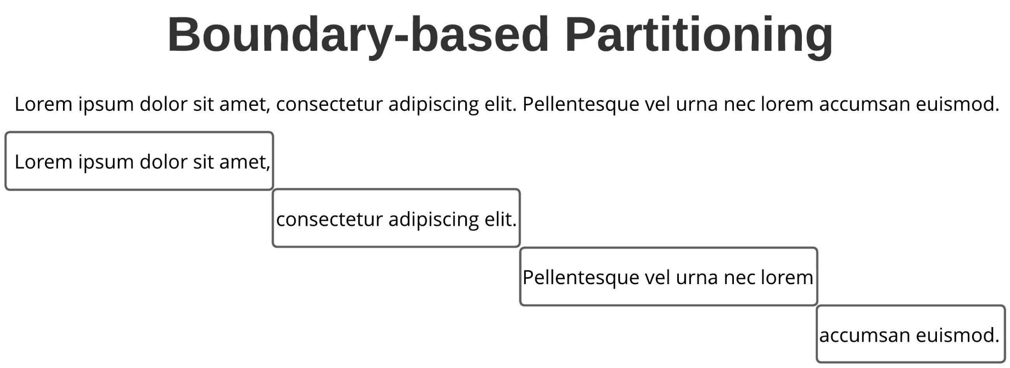

Figure 5: Example of Boundary-Based Partitioning ($m = 5, B(t) =$ sentence boundary)

4.3.4 Multiple Boundaries Partitioning

Multiple Boundaries Partitioning, as represented in Figure 6, allows for multiple boundaries within a partition but ensures that every partition ends at a boundary and the size of each partition is maximized, keeping partition sizes close to the desired threshold m.

4.3.5 Hierarchical Partitioning

In structured texts, such as academic articles, the content is often organized hierarchically. This hierarchy can range from the broadest divisions, like chapters or sections, to finer-grained subdivisions like subsections, paragraphs, or even individual sentences. Therefore, Hierarchical Partitioning, as

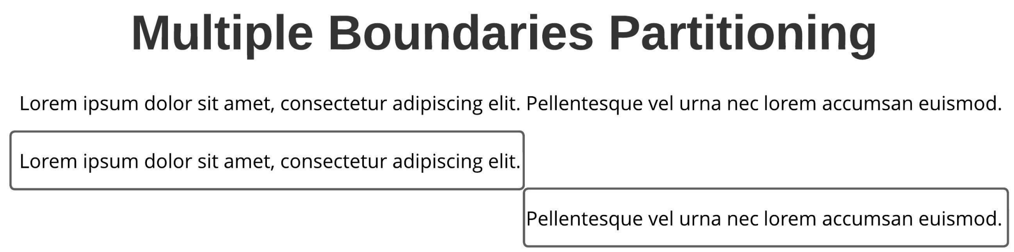

Figure 6: Example of Multiple Boundaries Partitioning ($m = 5, B(t) =$ sentence boundary)

represented in Figure 7, focuses on generating partitions at a specific level of the document's hierarchy. This ensures that the content within each partition is coherent and contextually linked.

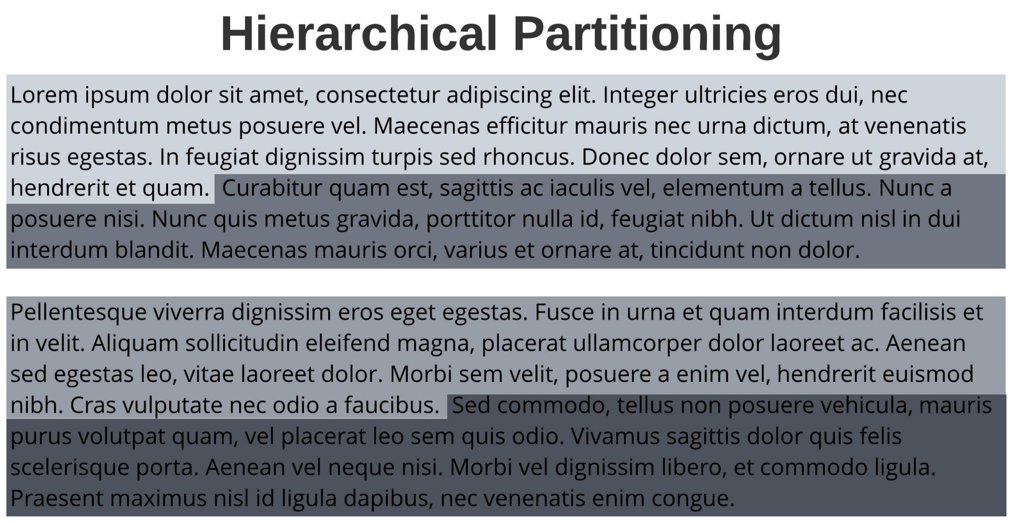

Figure 7: Example of Hierarchical Partitioning ($m = 50, L =$ paragraph and sentence)

4.4 Considerations for Partitioning

When partitioning text, several considerations come into play:

- **Overlap**: To ensure continuity, segments can have overlapping tokens. This ensures that the LLM doesn't lose contextual information at the boundaries.

Algorithm 7: Multiple Boundaries Partitioning

Data: Text T as sequence of tokens $(t_1, t_2, \ldots, t_n)$, Maximum partition size m, Boundary function $B(t)$

Result: List of partitions P

```
begin
    Initialize an empty list P;
    Initialize an empty partition p;
    for each token t_i in T do
        Append t_i to p;
        if Boundary function B(t_i) is true then
            if |p| ≥ m OR the next token t_{i+1} makes |p| > m then
                Append p to P;
                Reset p to an empty partition;
    if p is not empty then
        Append p to P;
    return P;
```

Algorithm 8: Hierarchical Partitioning

Data: Text T with hierarchical structure, Maximum partition size m, Desired hierarchy level L

Result: List of partitions P

```
begin
    Initialize an empty list P;
    Extract all units at level L from T as units;
    for each unit in units do
        if size of unit ≤ m then
            Append unit to P;
        else
            Subdivide unit into smaller units based on the next
              hierarchical level;
            Append each subdivided unit to P;
    return P;
```

- **Balance**: It's essential to strike a balance between having too many small segments and too few large ones. Too many small segments might increase the processing time, while too few large ones risk missing out on important details.
- **Order**: The order of segments can influence the results, especially when processing them sequentially. It's crucial to maintain the original order unless there's a specific reason to shuffle them.

5 Processing Strategies

5.1 Definition and Importance

After partitioning the text, the next crucial step is to utilize these partitions for various tasks. The nature of the task determines the processing strategy for these partitions. The tasks can be broadly classified into three categories based on their treatment of the partitions:

1. **Viewing the corpus as an aggregate:** Some tasks necessitate a comprehensive view of the entire corpus to derive meaningful outcomes. This is because the holistic context of the data provides crucial information.
    - *Example: Summarization* - When summarizing a document, it's essential to understand the entire narrative to produce a concise yet comprehensive summary. If one were to summarize individual partitions separately, the overarching themes or narratives might be lost.
    - *Example: Topic Modeling* - To discern the dominant topics in a collection of texts, the entire corpus must be considered to accurately capture recurring themes and subjects.
2. **Segregating the corpus:** Other tasks focus on specific segments of the text, allowing them to be processed independently. This independence arises from the task's inherent nature, where global context might not add significant value.
    - *Example: Named Entity Recognition (NER)* - The goal here is to identify entities like names, places, and organizations in the text. Each segment can be processed in isolation since the identification of an entity in one segment is typically independent of another.

- *Example: Part-of-Speech Tagging* - Assigning grammatical categories (like noun, verb, adjective) to individual words can be achieved without requiring the context of the entire document.

3. **Aggregating results from separated texts:** This category pertains to tasks that, while treating the partitions independently, require a subsequent aggregation phase to consolidate or compare results.

    - *Example: Sentiment Analysis* - If one wants to gauge the overall sentiment of multiple reviews, each review (or partition) can be analyzed separately. However, a subsequent aggregation step is essential to derive an overall sentiment score.
    - *Example: Document Clustering* - Individual documents or partitions might be represented in a vector space independently, but clustering them requires considering all the vectors together to discern patterns or groups.

Recognizing the classification of a task is pivotal, as it informs the preprocessing, partitioning, and post-processing strategies to be employed. Tailoring the approach based on task classification ensures efficient processing while preserving the integrity and intent of the data.

5.2 Mathematical Formulation

Let T_i represent a partitioned subtext from the original text T. Our objective is to perform independent processing on each subtext T_i and subsequently combine the outputs from all windows to generate a final result, contingent on the task's requirements.

Mathematically, the problem can be stated as:

$$\begin{aligned}
&\text{Independent Processing of Subtexts:}\\
&P(T_i) = R_i \quad \forall i \in \{1, 2, \ldots, k\}\\
&\text{Combination of Outputs:}\\
&R_c = \text{Combine}(R_1, R_2, \ldots, R_k)
\end{aligned}$$

Where T_i represents the partitioned subtext, $P(T_i)$ represents the independent processing function applied to subtext T_i, resulting in output R_i and R_c represents the combined output, which is generated by combining the outputs from all subtexts.

5.3 Strategies for Processing

5.3.1 Stuffing

Stuffing, as represented in Figure 8, is a streamlined approach to text processing within LLMs. It involves the straightforward insertion of a list of documents into a single prompt, which is then passed on to the LLM for analysis. This technique is particularly well-suited for scenarios where documents are relatively small in size, and where a single API call can efficiently handle the data. However, it's essential to note that the effectiveness of this technique may be limited by the content length restrictions imposed by the LLM, making it most suitable for situations where document size and quantity align with these constraints.

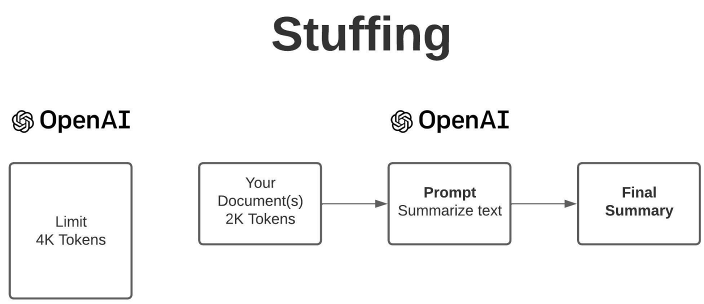

Figure 8: Example of Stuffing (Summarization). Adapted from: https://www.youtube.com/watch?v=f9_BWhCI4Zo

Algorithm 9: Stuffing

Data: Partitions $P_1, P_2, \ldots, P_N$
Result: Aggregated result R
begin
  $R \leftarrow \text{StuffFunction}(P_1, \ldots, P_N)$;
  return R;

5.3.2 Map Reduce

The Map-Reduce strategy, originally proposed for processing large datasets in parallel across distributed clusters, can be adapted for text processing tasks in LLMs (Dean et al., 2008). It consists of two primary steps. Initially, each partition is processed independently using a 'map' function, transforming the partition into a desired intermediate form. Following this, a 'reduce' function aggregates these intermediate results to produce a consolidated output. In this way, Map-Reduce allows for parallel execution, improving efficiency and reducing processing time. Additionally, it can extract specific information from each document to contribute to a more comprehensive final result. This strategy is particularly effective for tasks like summarization, where individual segment summaries can be further synthesized into a comprehensive summary as represented in Figure 9.

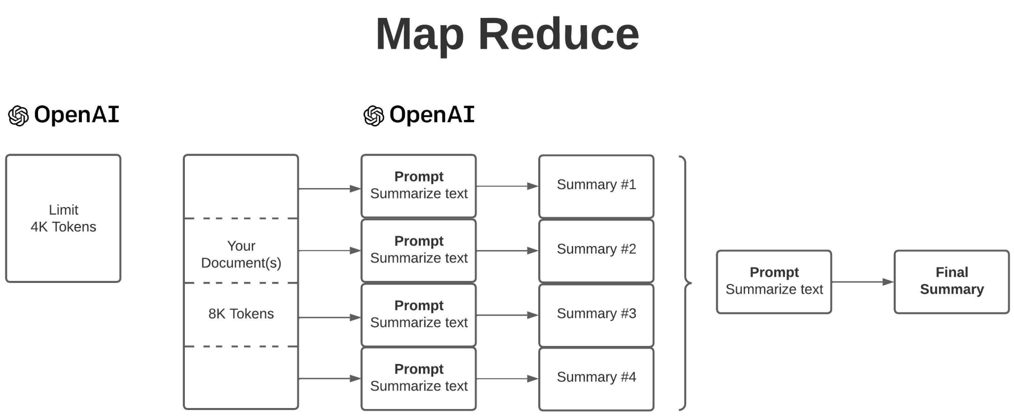

Figure 9: Example of Map Reduce (Summarization). Adapted from: https://www.youtube.com/watch?v=f9_BWhCI4Zo

Algorithm 10: Map-Reduce

Data: Partitions $P_1, P_2, \ldots, P_N$
Result: Aggregated result R
begin
  for *each partition* P_i **do**
    $M_i \leftarrow \text{MapFunction}(P_i)$;
  $R \leftarrow \text{ReduceFunction}(M_1, \ldots, M_N)$;
  return R;

5.3.3 Elimination Processing

Elimination Processing, as represented in Figure 10, is a variation of the Map-Reduce technique, which introduces a distinctive approach in the 'reduce' phase. In contrast to the conventional Map-Reduce method, where intermediate forms generated by the ‘map’ function are combined all at once, Elimination Processing adopts a gradual approach to aggregation, combining the intermediate results in groups of n where n is a predefined numerical parameter. This approach offers its own set of advantages, particularly in scenarios where fine-grained control over the aggregation process is desirable. By breaking down the reduction into smaller, manageable steps, Elimination Processing provides enhanced flexibility and adaptability, making it particularly valuable for text processing tasks that require nuanced and controlled consolidation of information. However, this incremental approach may introduce additional computational complexity, potentially impacting processing efficiency and time, depending on the chosen value of n.

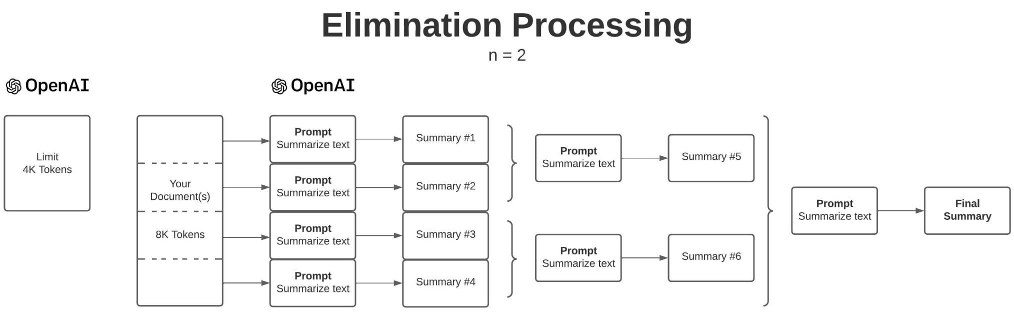

Figure 10: Example of Elimination Processing (Summarization). Adapted from: https://www.youtube.com/watch?v=f9_BWhCI4Zo

---

Algorithm 11: Elimination Processing

Data: Partitions $P_1, P_2, \ldots, P_N$, Reduce step n
Result: Aggregated result R

begin
  for *each partition* P_i **do**
    $M_i \leftarrow \text{MapFunction}(P_i)$;
  for $i = 0$ **to** $len(M)$ *by* n **do**
    $R \leftarrow \text{ReduceFunction}(M_i, \ldots, M_{i+n})$;
    add R to M;
  return R;

5.3.4 Sliding Window Processing

Sliding Window Processing, as represented in Figure 11, is an innovative adaptation of the Map-Reduce strategy, which has inspiration from the sliding window partitioning method. In this variant, the fundamental distinction lies in the 'reduce' stage. It operates by combining intermediate responses within overlapping segments, each of a fixed size w. The starting point of each segment is shifted by a constant step size s, defining the degree of overlap between consecutive segments. Sliding Window Processing excels in scenarios where maintaining contextual coherence throughout the processing pipeline is critical, making it particularly suitable for text processing tasks that benefit from incremental, fine-grained aggregation, such as dynamic summarization and context-aware document analysis.

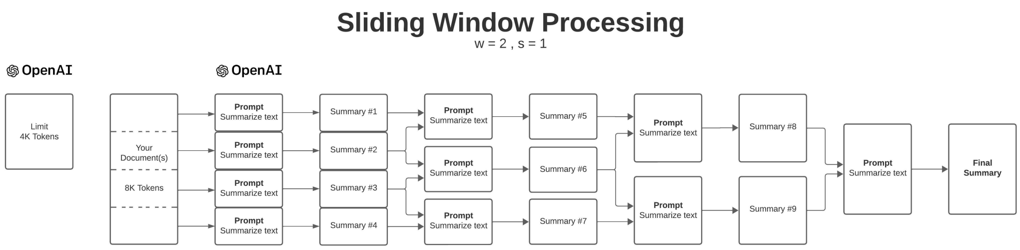

Figure 11: Example of Sliding Window Processing (Summarization). Adapted from: https://www.youtube.com/watch?v=f9_BWhCI4Zo

---

Algorithm 12: Sliding Window Processing

Data: Partitions $P_1, P_2, \ldots, P_N$, Window size w, Step size s
Result: Aggregated result R

begin
  for *each partition* P_i **do**
    $M_i \leftarrow \text{MapFunction}(P_i)$;
  for $i = 0$ **to** $len(M) - w + 1$ *by* s **do**
    $R \leftarrow \text{ReduceFunction}(M_i, \ldots, M_{i+w})$;
    add R to M;
  return R;

5.3.5 Map Re-Rank

Map Re-Rank, as represented in Figure 12, is a variation on map-reduce that introduces an innovative twist to document processing. It also consists of two primary steps. Initially, each partition is processed independently using a 'map' function, aiming not only to fulfill a specific task but also to provide a confidence score for the response. Following this, the 'reduce' function selects the response with the highest confidence score as the ultimate outcome. This approach shines when scalability is a primary concern, and it excels in scenarios where single-answer questions are paramount. However, it's worth noting that Map Re-Rank may not be the ideal choice when the need arises to synthesize information across multiple documents, as it primarily focuses on selecting the most confident response from individual documents.

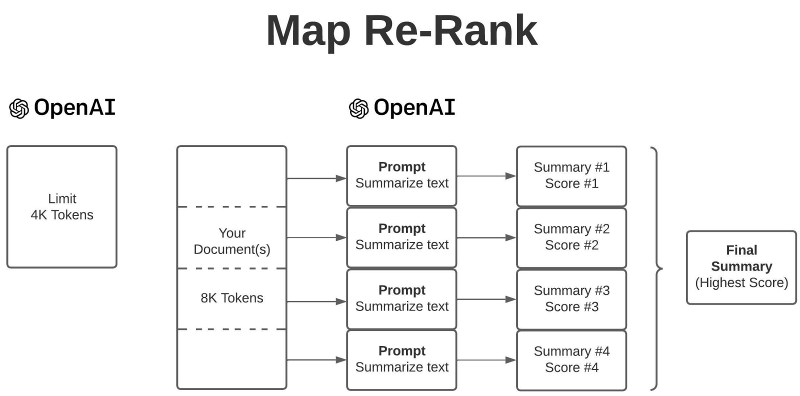

Figure 12: Example of Map Re-Rank (Summarization). Adapted from: https://www.youtube.com/watch?v=f9_BWhCI4Zo

---

Algorithm 13: Map Re-Rank

Data: Partitions $P_1, P_2, \ldots, P_N$
Result: Aggregated result R

begin
  for *each partition* P_i **do**
    $M_i, S_i \leftarrow \text{MapFunction}(P_i)$;
  $R \leftarrow M_i$ with highest S_i;
  return R;

5.3.6 Refine

Refine, as represented in Figure 13, is another approach to document processing that construct responses by iterating over input documents and continuously refining its answers. With each iteration, it feeds the current document in focus, and the latest intermediate answer into an LLM chain, resulting in a new answer. The advantages of Refine includes the capacity for continuous improvement and its ability to enhance precision over time. However, it comes with certain trade-offs, such as increased computational demands and extended processing durations due to the iterative nature of the process. The Refine chain excels in specific scenarios, such as long-form text generation and answer synthesis tasks.

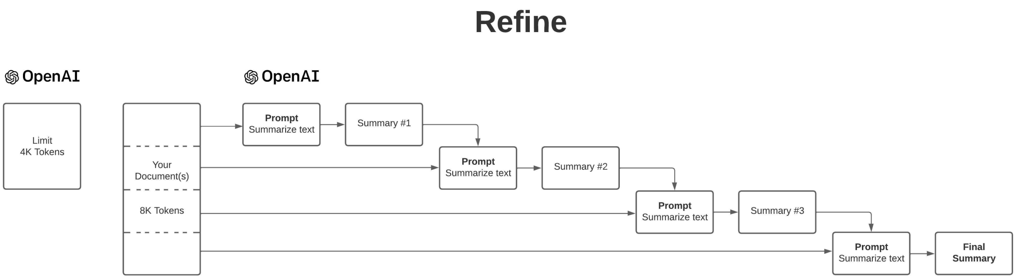

Figure 13: Example of Refine (Summarization). Adapted from: https://www.youtube.com/watch?v=f9_BWhCI4Zo

Algorithm 14: Refine

Data: Partitions $P_1, P_2, \ldots, P_N$
Result: Aggregated result R
begin
  for *each partition* P_i **do**
    $R \leftarrow \text{RefineFunction}(R, P_i)$;
  return R;

5.3.7 Weighted Refine

Weighted Refine, as represented in Figure 14, is an innovative approach to document processing and a variant of Refine method that introduces the use of weights during the response refinement stage. By incorporating weighted factors, this method places a higher significance on the intermediate responses, thereby influencing the formation of the new answer. This technique offers distinct advantages, including the ability to fine-tune and prioritize specific information, ultimately leading to more tailored and contextually relevant responses. However, it's worth noting that the introduction of weights also introduces complexity, which can result in increased computational demands and potentially longer processing durations. Weighted Refine excels in scenarios where precision and customization of answers are paramount, making it a valuable addition to the document processing toolkit.

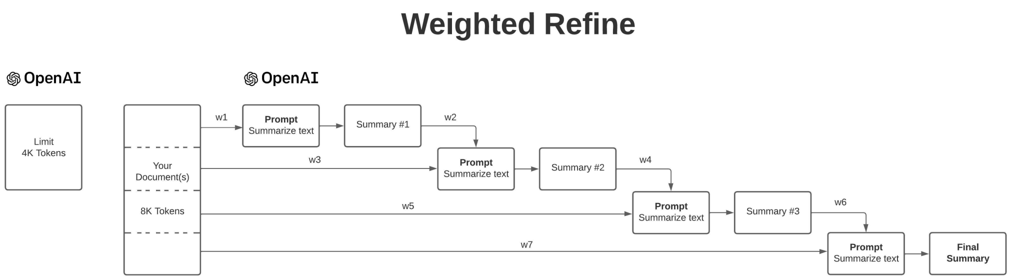

Figure 14: Example of Weighted Refine (Summarization). Adapted from: https://www.youtube.com/watch?v=f9_BWhCI4Zo

Algorithm 15: Weighted Refine

Data: Partitions $P_1, P_2, \ldots, P_N$
Result: Aggregated result R

```
begin
    for each partition P_i do
        R ← WeightedRefineFunction(R, P_i, w_1, w_2);
    return R;
```

5.4 Considerations for Processing

Efficient processing of text segments, especially within the context of LLMs, requires thoughtful considerations:

- **Model Selection**: Choose the appropriate LLM model for your specific document processing task. Different models may excel in tasks like summarization, translation, or sentiment analysis. Consider factors such as model size, language support, and pre-trained capabilities.

- **Model Fine-Tuning**: If available and applicable, consider fine-tuning the LLM on a domain-specific dataset or task. Fine-tuning can significantly improve model performance for your specific document processing needs.
- **Parallelism**: Utilize parallel processing when dealing with multiple segments. Distributing processing tasks across multiple cores or nodes can significantly improve performance.

6 Experiments on Metacritic Reviews

6.1 Dataset Description

The dataset for this experiment consists in Metacritic's user reviews of the game The Last of Us for Playstation 3. Each review has a date, score, textual review and a color grade indicating the sentiment: red (negative), yellow (neutral) and green (positive). In total, there were 14354 user reviews. However, it was only considered the reviews with text reviews and written in english. So, in total, there are 2341 user reviews.

6.2 Objective

Our goal is to generate a comprehensive summary of opinions for The Last of Us for Playstation 3. This task presents a unique challenge, as it requires to effectively harness the power of algorithms for sampling, partitioning and processing large text segments.

6.3 Experimental Setup

The experiments are designed to evaluate the combination of sampling methods, partitioning methods and subsequent processing techniques.

- **Sampling Methods**:
  - Uniform Sampling
  - Stratified Sampling with Known Labels (Ratings)
  - Stratified Sampling with Known Labels (Groups of Ratings)
- **Partitioning Methods**:
  - Simple Splitting

- Sliding Window
- Multiple Boundaries

- **Processing Methods**:
  - Map Reduce
  - Elimination Processing
  - Sliding Window Processing
  - Refine
  - Weighted Refine

6.4 Results

The methods and techniques mixed and its results were:

- **Uniform Sampling, Simple Splitting, Map-Reduce**:

  The Last of Us is a critically acclaimed video game praised for its emotional narrative, intense gameplay, and impressive visuals. It follows two characters, Joel and Ellie, in a post-apocalyptic world and features unique character development and environment-based combat. It is praised for its strong soundtrack and voice acting, and many consider it to be the best game of all time. However, some have found it to be too slow-paced and overrated. It is a difficult game with occasional frame rate drops, but is still an enjoyable experience that should not be missed.

- **Uniform Sampling, Simple Splitting, Eliminating Processing**:

  The Last of Us is a highly praised video game critically acclaimed for its engaging combat, stunning visuals, and immersive multiplayer mode. It is praised for its emotionally charged story, believable characters, and engaging gameplay, as well as its immersive world, intense characters, and realistic graphics and sound. Despite some criticism regarding AI, button mashing, and linear story, the game is seen as a masterpiece for its beautiful graphics, intense combat, and emotionally gripping narrative.

- **Uniform Sampling, Simple Splitting, Sliding Window Processing**:

  The Last of Us has been praised for its emotional story, engaging gameplay, and realistic graphics, but there are some issues such as the hand to hand combat, lack of music, and immersion-breaking elements. Despite its flaws, many consider it one of the best games of all time and a must-play for all gamers.

- **Uniform Sampling, Simple Splitting, Refine**:

  The Last of Us for the PS3 is widely acclaimed as one of the best games of its generation, praised for its emotional story, engaging characters, and intense gameplay. The graphics, sound, voice acting, and multiplayer mode are all praised, though there are some flaws such as a lack of replayability and some bugs. Players have a wide range of opinions, with some finding the experience to be gripping and emotionally resonant, while others find the story cliche or the gameplay tedious. However, most agree that it is a masterpiece and well worth playing.

- **Uniform Sampling, Simple Splitting, Weighted Refine**:

  The Last of Us is a critically acclaimed survival horror game that has mixed reviews but is generally praised for its storytelling, visuals, and unique gameplay. It is seen as one of the best games of all time by some, and is highly recommended by most. It has an engrossing story, compelling characters, great gameplay and atmosphere, and is considered a masterpiece by many. However, some criticize its AI and ending.

- **Uniform Sampling, Sliding Window, Map-Reduce**:

  The Last of Us is a critically acclaimed masterpiece from Naughty Dog that is praised for its emotionally charged story, thrilling and responsive gameplay, and stunning graphics. Players take the role of Joel and Ellie as they traverse a broken America, and the realistic violence adds to the game's dark and desperate tone. Multiplayer is a little bare, but the single player campaign provides an intense and emotionally draining experience. Despite some minor flaws, such as AI issues and some bugs, the overall package is well worth the play.

- **Uniform Sampling, Sliding Window, Eliminating Processing**:

  The Last of Us is a critically acclaimed masterpiece of a game that has been praised for its visuals, gripping story, believable characters, and intense gameplay. It is a unique experience that successfully blurs the lines between movies and games and is sure to be remembered for years to come.

- **Uniform Sampling, Sliding Window, Sliding Window Processing**:

The Last of Us is a critically acclaimed game that is praised for its outstanding graphics, emotionally engaging story, deep characters, intense gameplay, and replay value. It is an unforgettable experience that is worth the purchase despite its flaws, and is enjoyable for both single-player and multiplayer. It requires skill and strategy and is a must-play game that offers a rewarding and immersive experience. It has realistic brutality, but is not a shooting gallery and requires caution. It leaves the possibility for a sequel, though its ending may be unsatisfactory.

- **Uniform Sampling, Sliding Window, Refine**:

  The Last of Us is an acclaimed title praised for its visuals, soundtrack, and atmosphere. It features genre-defining mechanics, an emotionally engaging story and a unique relationship between Joel and Ellie. Despite occasional AI bugs, it is considered one of the best games of the generation and is highly recommended. Critics have praised the game for its story, atmosphere, and audio, as well as its unique action and emotion. The graphics and level design are praised, although there may be some issues with the gunplay and repetitive nature of the game. Ultimately, The Last of Us is an unforgettable experience that will leave a lasting impression.

- **Uniform Sampling, Sliding Window, Weighted Refine**:

  The Last of Us is an acclaimed action game developed by Naughty Dog that has been praised for its beautiful graphics, story, engaging characters, and challenging gameplay. It is widely considered one of the best games ever made and is praised for its emotional connection with the characters, its tension-filled gameplay, and its stunning visuals. Though some argue that it has flaws such as its zombie-like enemies, it is still an overall must-play experience.

- **Uniform Sampling, Multiple Boundaries, Map-Reduce**:

  The Last of Us is an acclaimed post-apocalyptic video game from Naughty Dog featuring great graphics, sound, AI, and an immersive story. It has been praised for its cinematic story-telling, emotional characters, intense atmosphere, and intense horror elements. It has been rated as one of the best games of 2013 and has a strong replay value. It has been criticized for its one-dimensional characters and repetitive gameplay, but is still considered one of the best games ever made.

- **Uniform Sampling, Multiple Boundaries, Eliminating Processing**:

  The Last of Us is an acclaimed title from Naughty Dog that has been praised for its captivating story, excellent voice acting, stunning visuals, and immersive sound design. It has won numerous awards and is widely considered one of the best PlayStation games of all time. It features realistic and lovable characters, and a gripping and believable story, however the gameplay and friendly AI could use some innovation. Despite some minor flaws, it is an unforgettable experience that will leave players feeling powerful emotions and never lose hope.

- **Uniform Sampling, Multiple Boundaries, Sliding Window Processing**:

  The Last of Us is an acclaimed game from Naughty Dog, praised for its amazing story, visuals, soundtrack, characters and gameplay. It combines adventure, survival, action, stealth and exploration, with great graphics and an emotional story. Despite some feeling it is overrated and not revolutionary, it is still considered one of the greatest games of all time and is a must-play for gamers.

- **Uniform Sampling, Multiple Boundaries, Refine**:

  The Last of Us is a critically acclaimed post-apocalyptic game that features an excellent story, realistic combat, stunning visuals, and an emotional soundtrack. Players join Joel and Ellie on an emotional journey as they navigate a dangerous post-apocalyptic world. Although the plot may be a bit underwhelming, the characters and story are captivating and engaging. The game is praised for its immersive gameplay, sound design, and deep themes. Despite a strange ending, The Last of Us is a must-play for fans of the genre.

- **Uniform Sampling, Multiple Boundaries, Weighted Refine**:

  The Last of Us is an acclaimed post-apocalyptic game by Naughty Dog that features stunning visuals, an intense and realistic combat system, an emotionally captivating story, and an immersive and captivating gaming experience. Critics have argued that the game doesn't quite live up to the immense hype it has received, but many agree it is one of the best games of all time. It incorporates a great story telling and top notch gameplay with a light crafting system. It also has a unique multiplayer experience with a faction system that adds to the replayability. The Last of Us delivers an unforgettable experience with an intense, emotional, and engaging story that

- **Stratified Sampling (Ratings), Simple Splitting, Map-Reduce**:

  The Last of Us is a critically acclaimed video game from Naughty Dog that has won the hearts of many players. It has stunning graphics, immersive story-telling, intense combat, crafting elements, and an awesome cast of characters. Despite some flaws such as clunky AI and repetitive gameplay, the positives of the game far outweigh the negatives. It is a must-have for any PS3 owner and is highly recommended for gamers of all levels.

- **Stratified Sampling (Ratings), Simple Splitting, Eliminating Processing**:

  The Last of Us is a critically acclaimed game that is considered one of the best of all time. It features stunning visuals, a deep and emotionally gripping story, great controls, and intense multiplayer. It has some minor bugs and glitches, but nothing that detracts from the experience. It is a must-own game for any Playstation 3 owner.

- **Stratified Sampling (Ratings), Simple Splitting, Sliding Window Processing**:

  The Last of Us is widely regarded as one of the best games of this generation, praised for its stunning graphics, narrative, performances, and gameplay. Despite some flaws such as unpredictable AI and too-easy supply acquisition, the overall experience is highly immersive and emotionally powerful, making it a must-play for any fan of video games.

- **Stratified Sampling (Ratings), Simple Splitting, Refine**:

  The Last of Us is universally acclaimed for its incredibly immersive story, characters, graphics, sound, and animations. It features intense combat and offers a unique post-apocalyptic experience that draws players in and never lets go. It is praised for its depth and emotional connection to the characters, with some claiming it to be one of the best games ever made. Despite some AI and difficulty issues, it is still a highly enjoyable game that stays with players long after the credits roll.

- **Stratified Sampling (Ratings), Simple Splitting, Weighted Refine**:

  The Last Of Us is highly praised for its stunning visuals, emotionally moving story, unpredictable plot, and combat system. The multiplayer was seen as comparatively poor to the single-player. It is widely considered to be the best game of the generation, and highly recommended for gamers of all types. It has earned many awards and is praised for its immersive story, characters, and atmosphere. It features unique gameplay with limited resources and no health regeneration, and has some graphical glitches and AI issues. Overall, it is an unforgettable experience that is praised for its story and gameplay.

- **Stratified Sampling (Ratings), Sliding Window, Map-Reduce**:

  The Last of Us is a critically acclaimed masterpiece from Naughty Dog that features stunning graphics, realistic sound effects, great voice acting, and an emotionally charged storyline. It immerses players into a post apocalyptic world filled with danger and suspense, requiring strategic thinking and improvisation to survive. With its replayability, script, cinematography, and horror elements, it has revived interest in the PS3 and consoles in general. It has bugs and the AI is not perfect, but it is a must-play game for mature gamers.

- **Stratified Sampling (Ratings), Sliding Window, Eliminating Processing**:

  The Last of Us is a critically acclaimed post-apocalyptic game from Naughty Dog that features an emotionally charged story, captivating gameplay, stunning visuals, and intense atmosphere. Its crafting system and multiplayer mode emphasize teamwork and survival, though there can be inconsistencies with the AI and some repetitiveness. Despite this, it is a great game and a must-play experience for any mature gamer.

- **Stratified Sampling (Ratings), Sliding Window, Sliding Window Processing**:

  The Last of Us is a highly praised game that has earned widespread acclaim for its emotionally engaging story, attractive visuals, and intense combat. Despite some criticisms regarding AI and puzzles, the game is generally seen as a masterpiece and has become a favorite amongst gamers of all ages and backgrounds.

- **Stratified Sampling (Ratings), Sliding Window, Refine**:

  The Last of Us is a critically acclaimed game praised for its stunning visuals, atmospheric music, and emotional story. It follows Joel and Ellie as they travel through a post-apocalyptic world, and although there may be some repetitive elements and difficult AI, the game is immersive and engrossing. It has been lauded for its beautiful visuals, emotional story, and gripping gameplay.

- **Stratified Sampling (Ratings), Sliding Window, Weighted Refine**:

  The Last of Us from Naughty Dog is a critically acclaimed game that has earned praise for its engaging story, visuals, characters, and gameplay. Players appreciate the game despite some issues with linearity, repetitive combat, and AI. It has been described as an emotionally charged experience that immerses players in a post-apocalyptic world and encourages them to question their morality. Many reviewers highly recommend it to anyone looking for a game that will leave a lasting impression. It has been praised for its production values, immersive world, and emotional narrative. It is widely considered to be one of the best games of all time.

- **Stratified Sampling (Ratings), Multiple Boundaries, Map-Reduce**:

  The Last of Us is a genre-defining third-person shooter game with stealth elements from Naughty Dog that has received universal praise for its visuals, characters, story, sound design, and gameplay. It follows Joel and Ellie, who must survive in a post-apocalyptic world after a virus. The reviews praised the game for its immersive and emotional story, realistic gunplay and stealth, crafting, and a fun multiplayer mode. The game has become a testament to the debate over whether video games can be considered art, with its attention to detail, impressive visuals, and emotionally engaging performances from its actors. The Last of Us

- **Stratified Sampling (Ratings), Multiple Boundaries, Eliminating Processing**:

  The Last of Us is generally praised for its immersive presentation, captivating story, and tactical gameplay. While there are some criticisms of its repetitive game-play, poor AI, and unsatisfying ending, it is still considered a great game overall and is compared to some of the greatest games of all time.

- **Stratified Sampling (Ratings), Multiple Boundaries, Sliding Window Processing**:

  The Last of Us is a highly praised game, seen by many as a masterpiece with an emotional story, immersive gameplay, and near-realistic graphics. It is a benchmark for single-player campaigns, with its unique multiplayer mode being both challenging and fun. Despite some criticism of its repetitive gameplay and cliched plot, it is almost universally praised and should not be missed.

- **Stratified Sampling (Ratings), Multiple Boundaries, Refine**:

  The Last of Us is a PS3 exclusive game highly praised for its cinematic storytelling, graphics, soundtrack, upgrade system, crafting and journal-using mechanics, and character development. There are criticisms of the game's repetitive gameplay and dumb AI, but it is still recommended for its emotional journey and beauty. It is a unique masterpiece that has made many players feel with the characters and kept them thrilled throughout the game. It is praised for its excellent audio, voice acting, and graphics, and is considered one of the greatest games of all time.

- **Stratified Sampling (Ratings), Multiple Boundaries, Weighted Refine**:

  The Last of Us is an acclaimed post-apocalyptic game praised for its story, characters, atmosphere, and graphics. It is unanimously rated as a must-play for PlayStation 3 players and features a mix of stealth and action gameplay, as well as an adventure element and multiplayer mode. It is considered one of the best games ever made for its emotional depth and compelling characters. Players have praised the game's visuals, plot-line, character development, and voice acting, while also noting the tough and realistic fighting. The Last of Us is deemed a masterpiece for its story, which features layers of messages and takeaways, and its characters,

- **Stratified Sampling (Groups of Ratings), Simple Splitting, Map-Reduce**:

  The Last of Us is widely praised as one of the best games of the PS3 era. It has stunning graphics, an emotionally gripping story, and intense gameplay. It features a crafting system and combat that takes some time to master, and the acting and character development are praised. There are some criticisms about its difficulty curve and lack of urgency, but overall, it is considered a classic and unforgettable experience.

- **Stratified Sampling (Groups of Ratings), Simple Splitting, Eliminating Processing**:

  The Last of Us is a highly praised game that has been lauded for its emotionally engaging narrative, stunning visuals, and tense and immersive gameplay. It has been considered a masterpiece of storytelling, visuals, and gameplay and is recommended for anyone looking for a great story and an immersive gaming experience. Although some have found the AI companions to be lacking and the story to be cliche, the game is still highly recommended and the multiplayer is enjoyable.

- **Stratified Sampling (Groups of Ratings), Simple Splitting, Sliding Window Processing**:

  The Last of Us from Naughty Dog is a highly praised post-apocalyptic and survival horror game that has been lauded for its cinematic storytelling, lifelike animations, creative gameplay, and replay value. It has been hailed as one of the best games of the seventh generation and the best game of 2013. Although some minor grievances such as AI interaction and lack of content have been noted, the game is still seen as perfect and is given a full 10/10 rating.

- **Stratified Sampling (Groups of Ratings), Simple Splitting, Refine**:

  The Last Of Us is hailed as a masterpiece due to its incredible story, graphics, and voice acting. It has won over 200 awards and is considered one of the greatest games of all time. Players are pulled into a post-apocalyptic world full of danger, and must use limited ammo to face intense and challenging scenarios. Despite some minor glitches and lack of character development, The Last Of Us is an all-time classic and a testament to the art form of video games.

- **Stratified Sampling (Groups of Ratings), Simple Splitting, Weighted Refine**:

  The Last of Us is widely acclaimed as one of the best video games of all time, praised for its emotionally gripping story, realistic graphics, and engaging atmosphere. While it has been criticized for its AI issues and linearity, many people still consider it to be an essential piece of interactive entertainment. It has earned near unanimous praise for its masterful narrative, beautiful graphics, and engaging gameplay. It is a unique experience that can be enjoyed by fans of all types of games.

- **Stratified Sampling (Groups of Ratings), Sliding Window, Map-Reduce**:

  The Last of Us is a critically acclaimed post-apocalyptic game that has become a beloved classic. It features stunning visuals, immersive gameplay, and an emotional story with characters that draw players in. The game has earned an impressive number of perfect scores and multiple awards, making it a must-play for any fan of the genre or gaming in general. It is praised for its character writing, setting, and relationship between Joel and Ellie, as well as its tactical and stealth-based gameplay and online mode. Despite its few criticisms, such as repetitive elements and predictable plot, it is highly recommended and considered to be one of the best

- **Stratified Sampling (Groups of Ratings), Sliding Window, Eliminating Processing**:

  The Last of Us is a critically acclaimed post-apocalyptic game with realistic gameplay, intense action sequences, a powerful story, and a great soundtrack. It is praised for its gripping story, beautiful visuals, immersive gameplay, and replayability, and is considered a must-play for any PlayStation 3 owners.

- **Stratified Sampling (Groups of Ratings), Sliding Window, Sliding Window Processing**:

  The Last of Us is a critically acclaimed video game developed by Naughty Dog that has earned praise for its engaging story, gripping plot, realistic characters, stunning visuals, and intense tactical combat. Although there are some issues with its third-person shooter gameplay, quicktime events, and generic and forgettable moments, the game is still highly regarded and is seen as one of the greatest games ever. It is a must-play for any fan of the genre.

- **Stratified Sampling (Groups of Ratings), Sliding Window, Refine**:

  The Last of Us is a critically acclaimed masterpiece from Naughty Dog that has been praised for its emotionally impactful story and characters, intense and realistic gameplay, award-winning voice acting, and gorgeous graphics. Fans of the game note some minor flaws such as a lack of weapon variety and some bugs, yet it is still considered one of the best video games ever created. With its heart-wrenching story and unforgettable gaming experience, The Last of Us is a must-play for any serious video game fan.

- **Stratified Sampling (Groups of Ratings), Sliding Window, Weighted Refine**:

  The Last of Us is a critically acclaimed game that has won universal 10/10 scores and is highly praised for its powerful story, great characters, satisfying gameplay, stunning graphics, and immersive atmosphere. Despite some criticisms of its linearity and repetitive elements, many consider it a masterpiece. The reviewer personally gave it a

9.7/10 rating, noting that while there are some minor issues, it is still an amazing game and is highly recommended.

- **Stratified Sampling (Groups of Ratings), Multiple Boundaries, Map-Reduce**:

  The Last of Us is a highly acclaimed survival horror game for Playstation 3, developed by Naughty Dog and released in 2013. It received widespread praise for its compelling story, intense atmosphere, and engaging gameplay. Players have praised the visuals, sound, and crafting mechanics, but some have criticized it for its overly scripted combat scenes and lack of exploration. Despite some flaws, it is considered one of the best games of all time and has been deemed a masterpiece by many.

- **Stratified Sampling (Groups of Ratings), Multiple Boundaries, Eliminating Processing**:

  The Last of Us is widely acclaimed as a modern masterpiece, praised for its emotional storytelling, memorable characters, beautiful graphics, and strategic gameplay. It is one of the best games of its generation, with immersive gameplay, stunning visual and sound design, and a unique combination of stealth, shooting, and resource management. Despite some minor bugs and glitches, it has been hailed as one of the best games of all time.

- **Stratified Sampling (Groups of Ratings), Multiple Boundaries, Sliding Window Processing**:

  The Last of Us is widely praised for its engaging story, visuals, voice acting, and immersive gameplay. While some reviewers have noted certain shortcomings, it is still considered an excellent game and a must-play for PlayStation owners. It features intense stealth/action gameplay, well-crafted characters, and a refined multiplayer mode.

- **Stratified Sampling (Groups of Ratings), Multiple Boundaries, Refine**:

  The Last of Us is widely praised for its story, characters, graphics, and immersive atmosphere. It is often considered one of the best games of the past generation and is a must-have for those wanting to experience an emotionally charged, cinematic story-driven game. It has some flaws such as linearity and AI, but its overall experience is worth it.

- **Stratified Sampling (Groups of Ratings), Multiple Boundaries, Weighted Refine**:

  The Last of Us has been widely praised for its story, characters, graphics, sound, cinematics, and replay value. Although some criticize the repetitive and unimaginative gameplay, it has been acclaimed for its creative ideas and classic gaming elements. It is often considered one of the best games of the generation and is remembered for its emotional storyline and ending. Naughty Dog's new IP from Neil Druckmann is a brutalist survival masterpiece, with incredible graphics, a great story and balanced stealth gameplay. It is highly recommended, and some even believe it is the best game ever made.

7 Conclusion

In this work, we have addressed the significant challenge of handling large contexts in LLMs like GPT and LAMMA. We have proposed a comprehensive set of strategies and techniques that go beyond the limits imposed by these models' context sizes.

We began by recognizing the inherent limitations of LLMs due to their fixed context size. This constraint hampers their ability to understand and generate text effectively when faced with larger documents or datasets. To overcome this limitation, we proposed and experimented with several strategies, including sampling, partitioning and processing methods.

In conclusion, our work opens up new avenues for leveraging the power of LLMs in dealing with extensive textual data. These strategies can find applications in natural language processing tasks, content generation, and data analysis, where a broader context is crucial for meaningful results. We anticipate that our findings will inspire further research and development in the field of language modeling and text analysis.

References

Beltagy, Iz, Matthew E Peters, and Arman Cohan (2020). "Longformer: The long-document transformer". In: *arXiv preprint arXiv:2004.05150* (cit. on p. 2).

Bird, Richard and Philip Wadler (1988). *An Introduction to Functional Programming*. Prentice Hall (cit. on p. 2).

Brown, Tom B, Benjamin Mann, Nick Ryder, Melvin Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. (2020). "Language Models are Few-Shot Learners". In: *arXiv preprint arXiv:2005.14165*. URL: https://arxiv.org/abs/2005.14165 (visited on 08/01/2023) (cit. on p. 1).

Conneau, Alexis, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov (2019). "Unsupervised cross-lingual representation learning at scale". In: *arXiv preprint arXiv:1911.02116* (cit. on p. 1).

Dean, Jeffrey and Sanjay Ghemawat (2008). "MapReduce: simplified data processing on large clusters". In: *Communications of the ACM*. Vol. 51. 1, pp. 107–113 (cit. on pp. 2, 16).

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova (2018). "BERT: Pre-training of deep bidirectional transformers for language understanding". In: *arXiv preprint arXiv:1810.04805* (cit. on p. 1).

Küppers, Anna, Ivan Vulić, Roi Reichart, and Anna Korhonen (2018). "Fragmenting and Reassembling Multilingual Documents". In: *arXiv preprint arXiv:1810.12504* (cit. on p. 2).

Liu, Yinhan, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov (2019). "RoBERTa: A robustly optimized BERT pretraining approach". In: *arXiv preprint arXiv:1907.11692* (cit. on p. 2).

OpenAI (2023). *GPT-4 Technical Report*. Tech. rep. OpenAI. URL: https://arxiv.org/pdf/2303.08774v3.pdf (visited on 08/02/2029) (cit. on p. 1).

Petroni, Fabio, Tim Rocktäschel, Sebastian Riedel, Patrick Lewis, Anton Bakhtin, Yuxiang Wu, and Alexander H Miller (2019). "Language Models as Knowledge Bases?" In: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pp. 2463–2473 (cit. on p. 1).

Radford, Alec, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever (2018a). "Improving Language Understanding by Generative Pretraining". In: URL: https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf (visited on 08/01/2023) (cit. on p. 1).

— (2018b). "Improving language understanding by generative pretraining". In: (cit. on p. 2).

Radford, Alec, Jeff Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever (2019). "Language Models are Unsupervised Multitask Learners". In: URL: https://d4mucfpksywv.cloudfront.net/better-language-models/language-models.pdf (visited on 08/01/2023) (cit. on p. 1).

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin (2017). "Attention is all you need". In: *Advances in neural information processing systems* 30 (cit. on p. 1).

Yang, Zhilin, Zihang Dai, Yiming Yang, Jaime G Carbonell, Ruslan Salakhutdinov, and Quoc V Le (2019). "XLNet: Generalized autoregressive pretraining for language understanding". In: *Advances in neural information processing systems* 32 (cit. on p. 1).

Zhang, Jingqing, Yao Zhao, Mohammad Saleh, and Peter J Liu (2019). "PEGASUS: Pre-training with extracted gap-sentences for abstractive summarization". In: *arXiv preprint arXiv:1912.08777* (cit. on p. 2).

Zhao, Rui, Bo Zhou, Yikang Yang, and Changyou Wu (2019). "Strategies for Training Large Vocabulary Neural Language Models". In: *arXiv preprint arXiv:1910.02079* (cit. on p. 2).