

---

***IS41720-Language Models and Methods-2024/25***

***Spring***

-

***Assignment One: NLP Task Analysis and Model Control***

---

- A. Introduction
- B. Methodology
- C. Results
- D. Discussion & Implications
- E. Conclusion

Submission Date : 02 April 2025

**Vaishnavi Kandikonda**

[vaishnavi.kandikonda@ucdconnect.ie](mailto:vaishnavi.kandikonda@ucdconnect.ie)

24216940

MSc. Information Systems 2024-25

University College Dublin

## Introduction

### **Chosen NLP Task:** Text Summarization

Text summarization is a very crucial and fundamental task in the field of Natural Language Processing (NLP), which addresses computer-human language interaction. This task is concerned with the process of reducing lengthy documents to shorter, more compact versions without sacrificing the intrinsic information contained in the original document. The significance of this task is greatly realized as it is heavily utilized in many scenarios, some of which are not limited to news aggregation, where current events are collected and summarized, research paper summarization for academic purposes, and content curation on several platforms. Being able to create summaries that are coherent and concise is of exceedingly great value, as it enables individuals and organizations to efficiently process, analyse, and consequently comprehend huge volumes of text data that would otherwise be too voluminous to process through. Summarization enables one to easily grasp main points in lengthy documents, making it a significant tool for researchers, professionals, and ordinary users.

### **Content Types:**

- **Technical Documents:** BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding – This type of content is dense with specialized terminology and complex information, posing unique challenges for summarization models(<https://arxiv.org/pdf/1810.04805> ).
- **News Articles:** Ukraine in maps: Tracking the war with Russia – News articles follow a structured format, emphasizing factual accuracy and clarity, contrasting with the complexity of technical documents.(<https://www.bbc.com/news/world-europe-60506682> )
- **Social Media Posts:** [R] Text-based backprop: Optimizing generative AI by backpropagating language model feedback : r/Machine Learning – Social media content is typically informal and varied in composition, further diversifying the dataset([https://www.reddit.com/r/MachineLearning/comments/1jn7jvg/r\\_text\\_based\\_backprop\\_optimizing\\_generative\\_ai\\_by/?rdt=36928](https://www.reddit.com/r/MachineLearning/comments/1jn7jvg/r_text_based_backprop_optimizing_generative_ai_by/?rdt=36928) ).

The specific types of content that were chosen for this research were chosen specifically with the goal of having a wide variety of text structures and levels of complexity. By doing this, this allows close and detailed examination of the ways in which different model control strategies can affect and influence summarization performance. Technical reports, for instance, are typically replete with dense and specialized information, presenting unique challenges to summarization. News articles, on the other hand, are designed to be presented in a structured way with factual information, providing a contrast with less formal styles of writing. In addition, social media posts are characteristic of being informal in style and mixed in composition, providing additional variety in the dataset. This careful variety in selected content types means that the experiments conducted span a wide variety of text types. This, in turn, allows meaningful conclusions to be drawn about the performance of the model across a variety of domains and uses.

**Experimental Strategy:** The set of experiments were carefully run using a range of various combinations of content types and different model controls. These involved changes in temperature settings and the use of particular system prompts that were crafted for the task in question. The main aim of this undertaking was to systematically record the outcomes derived from these experiments and examine the profound effect that these controls have on the summarization task in question. By intentionally modifying the temperature settings and using varying system prompts throughout the experiment, we sought to closely note and comprehend how these specific elements affect both the quality and unique attributes of the summaries produced by the model. This particular approach enables a deep and detailed understanding of the intricate relationship that exists between the model controls and the content types. As a result, it offers meaningful and revealing information that can be of great assistance in optimizing summarization tasks.

## Methodology

### Experimental Process:

1. **Content Selection:** Three documents were chosen, each from a different content type category. This selection ensured a diverse dataset that could highlight the effects of model controls across different types of text. The chosen documents were representative of their respective categories, providing a solid foundation for the experiments.

2. **Experimental Conditions:** For each content type, six experiments were performed:

- Low temperature (0.0 - 0.4)
- Low temperature + system prompt 1
- Low temperature + system prompt 2
- High temperature (0.6 - 1.0)
- High temperature + system prompt 1
- High temperature + system prompt 2

**Experimental Matrix:** The matrix includes detailed documentation of each experiment's conditions and outputs. Below is a summary of the matrix:

Test ID	Content Type	Temperature	System Prompt	Output Notes
A1	Technical Document	Low (0.3)	None	Concise but lacks contextual details
A2	Technical Document	Low (0.2)	Prompt 1	Slightly more structured, but still lacks depth
A3	Technical Document	Low (0.1)	Prompt 2	Too brief, missing key points
A4	Technical Document	High (0.8)	None	More expansive, but introduces slight redundancy
A5	Technical Document	High (1.0)	Prompt 1	Verbose but captures more key elements
A6	Technical Document	High (0.7)	Prompt 2	More detailed, but introduces slight hallucination
B1	News Article	Low (0.2)	None	Short and factual, but lacks nuance
B2	News Article	Low (0.4)	Prompt 1	Structured approach, slightly better than B1
B3	News Article	Low (0.3)	Prompt 2	Overly brief, key points omitted
B4	News Article	High (1.0)	None	More detailed but introduces slight opinionated tone

Test ID	Content Type	Temperature	System Prompt	Output Notes
B5	News Article	High (0.9)	Prompt 1	Balanced summary with minor redundancy
B6	News Article	High (0.6)	Prompt 2	More engaging but has minor factual drift
C1	Social Media Post	Low (0.1)	None	Captures the main sentiment, but is too literal
C2	Social Media Post	Low (0.3)	Prompt 1	Improves readability but loses some nuances
C3	Social Media Post	Low (0.4)	Prompt 2	Too generic
C4	Social Media Post	High (0.6)	None	More engaging summary, but occasionally exaggerates sentiment
C5	Social Media Post	High (0.7)	Prompt 1	Dynamic but adds unintended emphasis
C6	Social Media Post	High (0.9)	Prompt 2	Creatively reworded, but may misrepresent original intent

## Results

### Key Patterns:

#### Critical Patterns:

- Low Temperature:** Summaries, by the nature of their definition, need to be concise and to the point; however, this brevity may be at the cost of depth and contextual information that is vital to proper comprehension of the subject matter. For instance, in the specific instance of technical reports, low temperature settings returned summaries that were short and direct but tended to overlook useful contextual information that is essential to achieving a comprehensive level of understanding of the subject matter under discussion. This particular trend appeared to be consistent across various content categories, and hence the indication is that low temperatures prefer brevity to rich and detailed information.

- **High Temperature:** Under high temperature settings, the generated summaries are far more detailed and rich in their information. However, this increased detail also tends to bring redundancy or mild hallucinations into the text. High-temperature settings, in particular, generated summaries that were more detailed and included a larger range of details. However, there were instances where such summaries included redundant information or mild inaccuracies. This was most prominent in the instance of news articles, where using high-temperature settings generated summaries that, while detailed, tended to become redundant or reiterated points unnecessarily at times.

- **System Prompts:** The purpose of using structured prompts is to ensure the overall content is more readable and coherent, although there are instances where the use of this structured approach can inadvertently result in a loss of some of the subtle meanings or may introduce an unwanted stress on particular points. The utilization of system prompts has significantly assisted in presenting summaries in a more organized manner, hence rendering them not just more readable but more coherent for the reader as well. That said, it should be noted that such a gain in clarity and organization may at times be attained at the expense of a loss in subtle nuances or, in some instances, an unwanted stress on particular points within the content. For example, for social media posts, though these prompts have indeed rendered the overall content more readable, at times they have altered the original tone or sentiment conveyed by the author.

### **Screenshots:**

**1. Content Type :** Technical Document ; **Temperature:** Low (0.3) ; **System Prompt:** None

huggingface.co

Homepage - University College Dublin

IS41720 Zone - a Hugging Face Space by courtneyf2

Week 4: Controlling Model Behaviour   Week 5: NLP Tasks   Week 8: Error Detection   **Assignment 1**

### Assignment Testing & Evaluation

Select NLP Task

Summarisation

Enter Text

"Abstract  
We introduce a new language representation model called BERT, which stands for Bidirectional Encoder Representations from Transformers. Unlike recent language representation models (Peters et al., 2018a; Radford et al., 2018), BERT is designed to pre-train deep bidirectional representations from unlabeled text by jointly conditioning on both left and right context in all layers. As a result, the pre-trained BERT model can be fine-tuned with just one additional output layer to create state-of-the-art models for a wide range of tasks, such as question answering and language inference, without substantial task-specific architecture modifications.  
BERT is conceptually simple and empirically powerful. It obtains new state-of-the-art results on eleven natural language processing tasks, including pushing the GLUE score to 80.5% (7.7% point absolute improvement),

Upload a File (TXT or PDF)

Drop File Here  
or  
Click to Upload

Select Temperature

0.1 0.3 1

System Prompt

None: Default (no specific role)

**Generate Response**

Output

BERT is a language representation model that pretrains deep bidirectional representations from unlabeled text by considering both left and right context in all layers. It can be fine-tuned with minimal modifications to achieve state-of-the-art performance on various natural language processing tasks. BERT outperforms previous models on tasks like question answering and language inference. It uses a masked language model and next sentence prediction task for pre-training, enabling bidirectional representations. BERT reduces the need for task-specific architectures and advances the state of the art in eleven NLP tasks. The code and pre-trained models are available online.

## 2. Content Type : News Article ; Temperature: Low (0.2) ; System Prompt: None

huggingface.co

Homepage - University College Dublin

IS41720 Zone - a Hugging Face Space by courtneyf2

Spaces   courtneyf2 / IS41720\_Zone   like   Running   App   Files   Community

### Language Models & Methods Lab Interface

Week 4: Controlling Model Behaviour   Week 5: NLP Tasks   Week 8: Error Detection   **Assignment 1**

### Assignment Testing & Evaluation

Select NLP Task

Summarisation

Enter Text

Enter your text here...

Upload a File (TXT or PDF)

Ukraine in maps: Tracking the war with Russia.pdf   5.9 MB

Select Temperature

0.1 0.2 1

System Prompt

None: Default (no specific role)

**Generate Response**

Output

The paper introduces BERT (Bidirectional Encoder Representations from Transformers), a new language representation model that pretrains bidirectional representations from unlabeled text by conditioning on both left and right context in all layers. BERT can be fine-tuned with minimal task-specific modifications to achieve state-of-the-art performance on various natural language processing tasks.

Key points and arguments presented in the paper include:

1. Current language representation models are limited by unidirectionality, hindering their effectiveness for downstream tasks.
2. BERT addresses this limitation by utilizing a masked language model pre-training objective, allowing deep bidirectional Transformer pre-training.
3. BERT demonstrates the importance of bidirectional pre-training for language representations compared to unidirectional approaches used in previous models like OpenAI GPT and ELMo.
4. BERT reduces the need for complex, task-specific architectures by achieving state-of-the-art performance across multiple NLP tasks.
5. The paper showcases BERT's advancements by improving results on eleven NLP tasks, such as question answering and language inference.
6. The availability of code and pre-trained models for BERT is highlighted, making it accessible for further research and applications.

In conclusion, the paper emphasizes the significance of rich unsupervised pre-training for enhancing language understanding systems, particularly showcasing the benefits of deep bidirectional architectures like BERT across a wide range of NLP tasks.

### 3. Content Type : Social Media Post; Temperature: Low (0.1) ; System Prompt: None

Language Models & Methods Lab Interface

Week 4: Controlling Model Behaviour   Week 5: NLP Tasks   Week 8: Error Detection   **Assignment 1**

#### Assignment Testing & Evaluation

Select NLP Task  
Summarisation

Enter Text  
Enter your text here...

Upload a File (TXT or PDF)  
[R] Text based backprop: Optimizing generative AI by backpropagating langu... .pdf 497.7 KB

Select Temperature  
0.1

System Prompt  
None: Default (no specific role)

**Generate Response**

Output  
The text discusses a new framework called TextGrad that optimizes generative AI by backpropagating language model feedback. This framework leverages natural language feedback to critique and suggest improvements to AI systems, enabling automatic optimization across various tasks. TextGrad has been shown to be effective in solving science problems, optimizing treatment plans, designing molecules, coding, and optimizing agentic systems. The paper on this topic was published in Nature.

### 4. Content Type : Technical Document; Temperature: High (1.0); System Prompt: System Prompt 1

Language Models & Methods Lab Interface

Week 4: Controlling Model Behaviour   Week 5: NLP Tasks   Week 8: Error Detection   **Assignment 1**

#### Assignment Testing & Evaluation

Select NLP Task  
Summarisation

Enter Text  
"Abstract  
We introduce a new language representation model called BERT, which stands for Bidirectional Encoder Representations from Transformers. Unlike recent language representation models (Peters et al., 2018a; Radford et al., 2018), BERT is designed to pre-train deep bidirectional representations from unlabeled text by jointly conditioning on both left and right context in all layers. As a result, the pre-trained BERT model can be fine-tuned with just one additional output layer to create state-of-the-art models for a wide range of tasks, such as question answering and language inference, without substantial task-specific architecture modifications. BERT is conceptually simple and empirically powerful. It obtains new state-of-the-art results on eleven natural language processing tasks, including pushing the GLUE score to 80.5% (7.7% point absolute improvement),...

Upload a File (TXT or PDF)  
Drop File Here  
- Or -  
Click to Upload

Select Temperature  
1.0

System Prompt  
Prompt 1: Professional Analyst providing precise analysis

**Generate Response**

Output  
Summary:  
- BERT (Bidirectional Encoder Representations from Transformers) is introduced as a new language representation model designed to pretrain deep bidirectional representations from unlabeled text by considering both left and right contexts in all layers.  
- The pre-trained BERT model can be fine-tuned with minimal modifications to achieve state-of-the-art performance across various natural language processing tasks such as question answering and language inference.  
- BERT outperforms previous models on eleven NLP tasks, significantly improving scores on tasks like GLUE score, MultiNLI accuracy, SQuAD v1.1, and SQuAD v2.0 question answering tests.  
- The paper discusses the limitations of unidirectional language models for fine-tuning approaches and introduces BERT to address these constraints by using a "masked language model" (MLM) pre-training objective and a "next sentence prediction" task to pretrain text-pair representations.  
- BERT demonstrates the importance of bidirectional pretraining for language representations, reduces the need for task-specific architectures, and advances the state of the art in NLP tasks.  
- The availability of code and pre-trained models for BERT is highlighted, demonstrating its practical applicability and accessibility for further research and application in the NLP domain.



## 5. Content Type : News Article; Temperature: High (0.6) ; System Prompt: System Prompt 2

The screenshot shows the Hugging Face Language Models & Methods Lab Interface. The page is titled "Assignment Testing & Evaluation" and includes a navigation bar with links for Week 4, Week 5, Week 8, and Assignment 1. The interface is configured for a "Summarisation" task. The "Enter Text" field contains the text "Enter your text here...". The "Upload a File (TXT or PDF)" button is visible, and a file named "Ukraine in maps: Tracking the war with Russia.pdf" (5.9 MB) is uploaded. The "Select Temperature" slider is set to 0.6. The "System Prompt" dropdown is set to "Prompt 2: Creative Interpreter providing detailed insights". The "Generate Response" button is visible. The output field displays a detailed report by the BBC about the conflict between Ukraine and Russia, mentioning the relentless fighting in eastern Ukraine and the impact of Western support on Ukrainian resistance.

## 6. Content Type : Social Media Post ; Temperature: High (0.9) ; System Prompt: System Prompt 2

The screenshot shows the Hugging Face Language Models & Methods Lab Interface. The page is titled "Assignment Testing & Evaluation" and includes a navigation bar with links for Week 4, Week 5, Week 8, and Assignment 1. The interface is configured for a "Summarisation" task. The "Enter Text" field contains the text "Enter your text here...". The "Upload a File (TXT or PDF)" button is visible, and a file named "[R] Text based backprop: Optimizing generative AI by backpropagating langu... .pdf" (497.7 KB) is uploaded. The "Select Temperature" slider is set to 0.9. The "System Prompt" dropdown is set to "Prompt 2: Creative Interpreter providing detailed insights". The "Generate Response" button is visible. The output field displays a recent Reddit post about a groundbreaking paper on optimizing generative AI through text-based backpropagation, mentioning the advancement in artificial intelligence and the potential of this technique.

## Discussion & Implications

### Temperature Effects:

- **Higher Temperatures:** The influence of higher temperatures is associated with the generation of more comprehensive and detailed summaries with plenty of information. This increase in detail, however, at times brings with it redundancy, where the same details are repeated, or factual drift, which refers to small deviations from the accuracy of the facts presented. For example, for the case of news articles, it was observed that the use of high temperature values produced summaries that were very rich in detail, providing a better general overview. It was also observed, though, that such summaries at times contained redundant details or small factual inaccuracies relating to factual components. This observation highly suggests that while the use of a higher temperature can significantly enhance the level of detail captured, careful regulation of this parameter is required to prevent unnecessary redundancy from being introduced in the generated summaries.
- **Lower Temperatures:** At lower temperature levels, the output is a generation of very short summaries that, while effective in terms of brevity, tend to be poor in terms of depth and richness of important contextual information. For instance, within the framework of social media updates, the application of lower-temperature settings led to the generation of summaries that were, indeed, short and effectively conveyed the overall sentiment. However, these summaries tended not to encapsulate the subtleties and depth of detail inherent in the original update. This observation presents the notion that, while low temperatures might be useful for generating quick and easy overviews, they are not particularly well-suited to more complex and detailed information.

### System Prompts:

- **Structured Summaries:** System prompts are essential in organizing and structuring summaries. This, in turn, makes the text more readable to the reader. A good example is where a system prompt was utilized in organizing technical documents. The summaries produced were significantly more structured and, therefore, more readable and easy to comprehend. This greatly underscores the essential need to utilize prompts as a means of organizing information in a better way, thus enhancing readability to readers.
- **Loss of Nuances:** It has also been observed that prompts occasionally result in a loss of subtlety or create an unwanted emphasis. Particularly in social media updates, the utilization of system-generated prompts has, on some occasions, generated summaries that turned out

to be excessively generic, or which unintentionally highlighted particular points more than was originally intended. This indicates that, although prompts can undoubtedly enhance structural clarity and organization, they must be meticulously designed and considered in design in a way that will retain the original nuances and subtleties of the content.

### **Types of Content:**

- **Technical Documents:** Technical documents benefit immensely from the application of high temperatures alongside carefully structured prompts as the process yields the production of highly detailed and informative summaries. The application of high temperatures coupled with specific system prompts has been shown to generate summaries that not only detect more key elements but also provide a much more rounded view. The fact firmly implies that in handling technical documents, there exists an extremely critical imperative to achieve a fine balance between detailed information and structured guidance in order to develop genuinely effective and useful summaries.
- **News Articles:** In the case of news articles, careful balance must be struck between the temperature setting and the prompts used in order to successfully avoid any redundancy that would detract from the main message while at the same time maintaining factual accuracy. Using a moderate temperature setting with well-written system prompts, the result was a list of well-balanced summaries that were both extremely detailed and exactly accurate. This suggests that, in news articles, well-balanced treatment that smoothly blends rich detail with a clear and orderly structure is highly advantageous.
- **Social Media Posts:** Social media posts are best captured by the use of high temperatures and specially designed system prompts that aim to draw out engagement from the audience, but need to be handled cautiously in an attempt to avoid exaggeration of any sort. While we applied the use of high-temperature settings alongside carefully crafted system prompts, we observed that the resulting summaries not only were engaging but also succeeded in capturing the essence of the original posts. However, it must be observed here that, at times, such summaries tended to slightly exaggerate the post sentiment. This observation brings to light the imperative necessity for social media posts to strike a fine balance between inducing engagement and remaining faithful to their depiction.

### **Most Successful Combinations:**

- **Technical Documents:** Low temperature combined with the addition of structured prompts aimed at technical documents has been successful. This novel combination has been able to

generate summaries that not only conserve words but are also structured, which makes them by far more readable and comprehensible. The results from this approach are suggestive of technical documents, authored with a structured style aimed at clarity and conciseness, working to greatly improve understanding and ease of use on the reader's part.

- **Social Media Posts:** The combination of high temperature with prompts specifically designed for social media posts has resulted in dynamic and interesting summaries. These summaries were able to capture the essence of the original posts, presenting the content in an appealing way. This finding suggests that social media posts are greatly served by a dynamic style, one that is not only focused on engagement but also on readability to draw and hold the audience's attention.

### **Unintended Consequences:**

- **Hallucinations:** Rare hallucinations or exaggerated sentiment have taken place in the setting of high temperatures. High temperature settings in some instances have led to the production of summaries that contained information not initially present in the text, or summaries that exaggerated the sentiment to an extent not accurately representative of the sentiment. This indicates that one must deal cautiously and carefully with high temperatures in order not to introduce any inaccuracies that would distort the meaning of the original content.
- **Redundancy:** Redundancy in high-temperature environments. High-temperature environments occasionally led to summaries

## **Conclusion**

This study examined how model controls—prompts and temperature—affect text summarization. Lower temperatures gave brief, uninformative summaries, while higher temperatures provided engaging but repetitive output. System prompts structured content but occasionally altered subtleties, especially on social media.

Various content demanded different settings: technical reports demanded systematic prompts and medium temperatures for coherence, news articles demanded balanced controls for factuality, whereas social media postings flourished under high temperatures and personal prompts for entertainment.