**DAILY UPDATES**

**Day 1-3: Introduction**

Got familiarized with Natural Language Processing (NLP) and its applications in text processing tasks.

Understand the concepts of abstractive and extractive summarization techniques using the T5 transformer.

**Day 3-7-: Abstractive and Extractive Summarization**

Dived deeper into abstractive summarization: Generate concise and coherent summaries using T5 transformer's language model.

Explored extractive summarization: Select and extract relevant sentences from the source text to create summaries.

**Day 7-9: T5 Transformer**

Studied the architecture and working principles of the T5 transformer for various NLP tasks.

Examined the pre-trained T5 model and fine-tuning techniques for summarization tasks.

**Day 9-11: Paraphrasing with T5 Transformer and Split and Fetch Sentences**

Learnt about paraphrasing and its importance in NLP tasks.

Implemented paraphrasing using the T5 transformer to generate alternative versions of input sentences.

The "Split and Fetch Sentences" step involves dividing the given text into individual sentences and collecting only those sentences that are at least 15 characters long.

**Day 11-15: Keyword and Keyphrase Extraction using RAKE and PKE**

Had an understanding on RAKE (Rapid Automatic Keyword Extraction) algorithm for keyword and keyphrase extraction.

Implement RAKE to automatically extract important keywords from the text.

PKE (Python Keyphrase Extraction) is a library that provides tools and algorithms for automatically extracting keyphrases (important words) from a given text.

**Day 15-16: T5 for Question Generation and mapping and Word Sense Disambiguation**

Explored question generation using the T5 transformer.

Fine-tuned the T5 model to generate meaningful questions based on input statements.

And we learnt about Mapping which is the process of associating or linking elements from one set to corresponding elements in another set based on certain rules or criteria.

Word Sense Disambiguation (WSD) is the process of determining the correct sense or meaning of a word in a given context.

**Day 16-17: WordNet and Sense2Vec for Distractor Generation**

Introduced WordNet and Sense2Vec as resources for NLP tasks.

Learnt how to use WordNet and Sense2Vec to generate distractors for multiple-choice questions.

**Day 17-18 : Improvement with ConceptNet:**

To improve the quality of distractors for Multiple Choice Questions (MCQs), the code utilizes ConceptNet to find related words based on the "PartOf" relationship with the given word, thus enhancing the diversity and relevance of the generated options.

**Day 18- 20: Finding and Mapping Distractors to Keywords**

In the "Find and Map Distractors to Keywords" step, the code searches for incorrect options (distractors) related to each important word (keywords) and maps them to create Multiple Choice Questions (MCQs).

**Day 20-23: Integration and Practice**

Integrated the learned techniques into a comprehensive NLP pipeline.

Practiced using different datasets and texts to perform summarization, paraphrasing, keyword extraction, question generation, and distractor generation.

**Day 23-26: Evaluation and Fine-Tuning**

Discussed various evaluation metrics for each task to assess the quality of results.

Fine-tuned the models and algorithms based on evaluation feedback to enhance performance.

**Recap of what we have learnt:**

**Abstractive and Extractive Summarization using T5 Transformer:**

Abstractive Summarization: Generate concise and coherent summaries by paraphrasing and rephrasing the original text using the T5 transformer's language model.

Extractive Summarization: Select and extract important sentences from the source text to create summaries, without modifying the original sentences.

**About T5 Transformer:**

T5 Transformer: An advanced NLP model capable of performing various tasks, including summarization, paraphrasing, question generation, and more.

Pre-training and Fine-tuning: The process of training a language model on a large corpus of text and fine-tuning it on specific tasks to improve its performance.

**Paraphrasing with T5 Transformer:**

Paraphrasing: The task of rephrasing a sentence or text while preserving its original meaning.

T5 for Paraphrasing: Using the T5 transformer to generate alternative versions of input sentences.

**Keyword and Keyphrase Extraction using RAKE & PKE:**

RAKE Algorithm: A rapid and automatic keyword extraction technique that identifies important keywords and keyphrases from a given text.

PKE (Python Keyphrase Extraction) is a library that provides tools and algorithms for automatically extracting keyphrases (important words) from a given text.

**T5 for Question Generation:**

Question Generation: Creating meaningful questions based on input statements or paragraphs.

T5 for Questions: Utilizing the T5 transformer to generate questions from input text.

Mapping: The process of linking elements between two sets based on specified rules or criteria.

Word Sense Disambiguation (WSD): The process of identifying the correct meaning of a word in a specific

context.

**Distractor Generation using WordNet and Sense2Vec & Conceptnet:**

Distractor Generation: Generating incorrect or misleading options for multiple-choice questions.

WordNet: A lexical database that provides semantic relationships between words.

Sense2Vec: An extension of Word2Vec that embeds multi-sense word representations.

Conceptnet: ConceptNet is a knowledge graph that connects words and phrases of natural language with labeled edges.

**Integration and Practice:**

Integrating the learned techniques into a comprehensive NLP pipeline to perform summarization, paraphrasing, keyword extraction, question generation, and distractor generation.

**Evaluation and Fine-Tuning**

Evaluating the performance of NLP models using appropriate metrics for each task.

Fine-tuning the models based on evaluation feedback to improve their quality and accuracy.

**Real world application**

* Education and Learning: MCQs are commonly used in educational settings, such as schools, colleges, and online courses. They offer an efficient way to assess students' knowledge and understanding of a subject. MCQ-based assessments can cover a wide range of topics, providing educators with valuable insights into students' strengths and weaknesses.
* Standardized Tests: Many standardized tests, such as SAT, GRE, TOEFL, and various professional certification exams, utilize MCQs as part of their evaluation process. These tests help measure a candidate's aptitude, skills, and knowledge, making the assessment process fair and standardized.
* Recruitment and Hiring: In the corporate world, MCQs are often used as part of pre-employment tests during the recruitment process. Employers can evaluate candidates' abilities and knowledge relevant to the job requirements efficiently.
* Competitive Examinations: Government agencies and institutions often conduct competitive exams for admissions and job placements. MCQ-based exams simplify the evaluation process for a large number of applicants.
* Training and Assessment in the Workplace: MCQs are utilized in employee training and assessment programs. Companies can assess employees' understanding of important policies, procedures, and product knowledge.
* Online Assessments and Quizzes: Online platforms, e-learning portals, and MOOCs (Massive Open Online Courses) often use MCQs to provide interactive assessments and quizzes, enhancing the learning experience for learners worldwide.
* Medical and Healthcare: MCQs are used in medical exams for licensing and specialty certifications. They help evaluate medical students, doctors, and other healthcare professionals' knowledge and expertise.
* Market Research and Surveys: In market research, MCQs are employed to collect quantitative data from survey respondents. This method enables researchers to analyse and interpret responses efficiently.
* Content Evaluation: MCQs can be used to evaluate the effectiveness of educational content and training materials. By integrating MCQs into courses, instructors can gauge how well learners grasp the material.
* Automated Grading: MCQs can be easily graded by automated systems, saving time and effort for educators and institutions.
* Game-Based Learning: In the realm of educational games and gamified learning, MCQs provide interactive challenges and reward systems for learners.
* Overall, MCQs offer a versatile and scalable way to assess knowledge, understanding, and critical thinking skills. Their practicality in grading and efficiency in evaluation make them an essential tool in education, training, and professional assessments in the real world.

**The next step for further development is to integrate Bengali language support into the NLP techniques and tools we've covered in this 1 month internship. This involves adapting the existing models and algorithms to process and generate text in the Bengali language. Here's how it can be beneficial:**

* Increased Accessibility: Integrating Bengali language support enables NLP applications to reach a wider audience, especially in regions where Bengali is spoken and used extensively. This inclusivity can have a significant impact on education, information dissemination, and access to services.
* Empowering Local Businesses: By supporting Bengali language processing, local businesses and organizations in Bengali-speaking regions can benefit from NLP-driven tools, such as sentiment analysis for customer feedback, content summarization, and automated customer support.
* Educational Applications: Bengali language support in NLP can facilitate the development of e-learning platforms and educational resources in Bengali. This would make quality education more accessible to Bengali-speaking learners, aiding in their language proficiency and academic growth.
* Content Creation and Translation: NLP tools capable of handling Bengali text can assist in content creation, translation, and paraphrasing tasks, making it easier to generate content in Bengali or translate it from other languages into Bengali.
* Native Language Chatbots: With Bengali language support, chatbots and virtual assistants can communicate with users in their native language, enhancing user experience and ensuring better interactions.
* Sentiment Analysis and Social Media Monitoring: Analysing Bengali text from social media platforms and online discussions can provide valuable insights into public sentiment, allowing businesses and organizations to better understand their Bengali-speaking audience.
* Legal and Government Services: Government agencies and legal institutions can use NLP tools to process and analyse Bengali language documents, leading to more efficient document management and information retrieval.
* Research and Development: Researchers and developers in Bengali NLP can advance the field by creating language-specific datasets, fine-tuning existing models, and developing innovative applications tailored to Bengali linguistic nuances.
* Preserving Cultural Heritage: Bengali language support can play a crucial role in preserving the cultural heritage of Bengali-speaking communities, as it enables the digitization and analysis of historical texts and literature.
* Natural Disaster Response: In regions prone to natural disasters, NLP applications can help process Bengali text from social media posts and news sources to provide timely updates and support disaster response efforts.
* Incorporating Bengali language support into NLP techniques demonstrates a commitment to linguistic diversity, cultural appreciation, and social inclusion. It empowers Bengali-speaking communities, fosters economic growth, and opens up new opportunities for research, innovation, and global collaboration in the field of NLP.

**Working Status:**

Currently, we are actively working on translating our project into Bengali and integrating the latest version of the Llama (llama v2) tool into our internship report. The project aims to enhance accessibility and usability for Bengali-speaking users.