INFOSNIP
TAILORED NEWS SUMMARIZATION

"Transforming extensive news articles into flash summaries"

Gulnaaz Shaikh
Parth Keyur Gawande
Sarthak Bora
Vikashini Sabarinathan

CONTENTS

- Introduction
- Problem statement
- Dataset Description
- Methodology
- Types of Summarization
- Evaluation metrics
- Comparative Study
- Conclusion
- Future Scope

INTRODUCTION

In our daily lives, we often contend with busy schedules that demand juggling various tasks, commitments, and responsibilities. It would be so hard for the people who are in a time crunch to read the news articles. InfoSnip provides a practical solution to the growing problem of information overload by condensing lengthy news articles into easily digestible summaries.
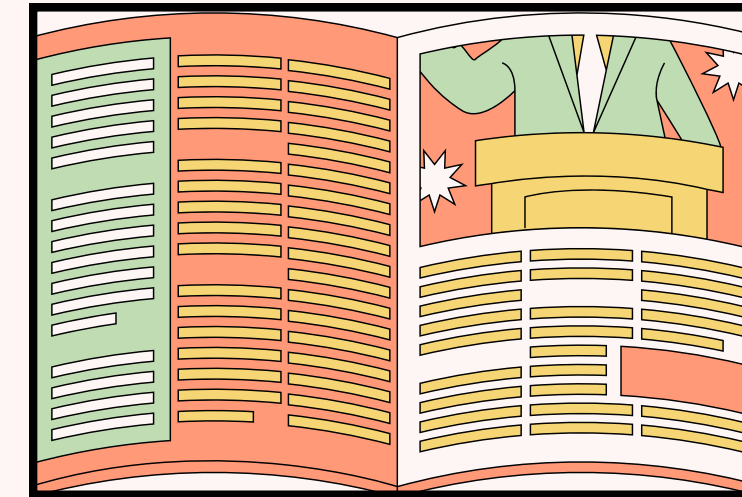
PROBLEM STATEMENT

In an era of information overload, people struggle to keep up with the vast amount of news articles available online.
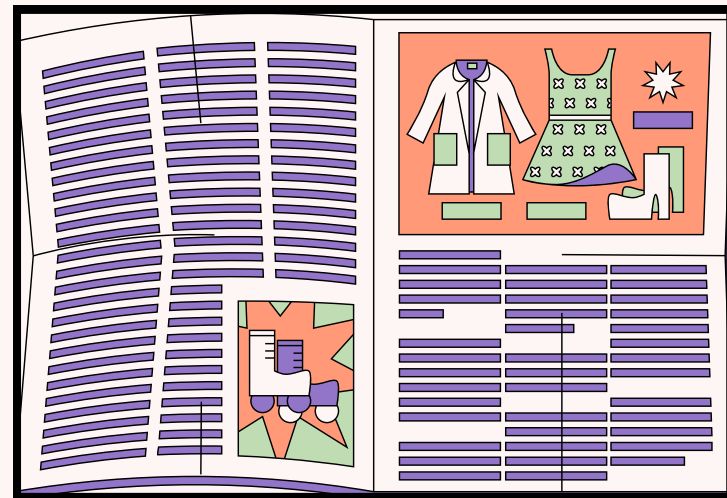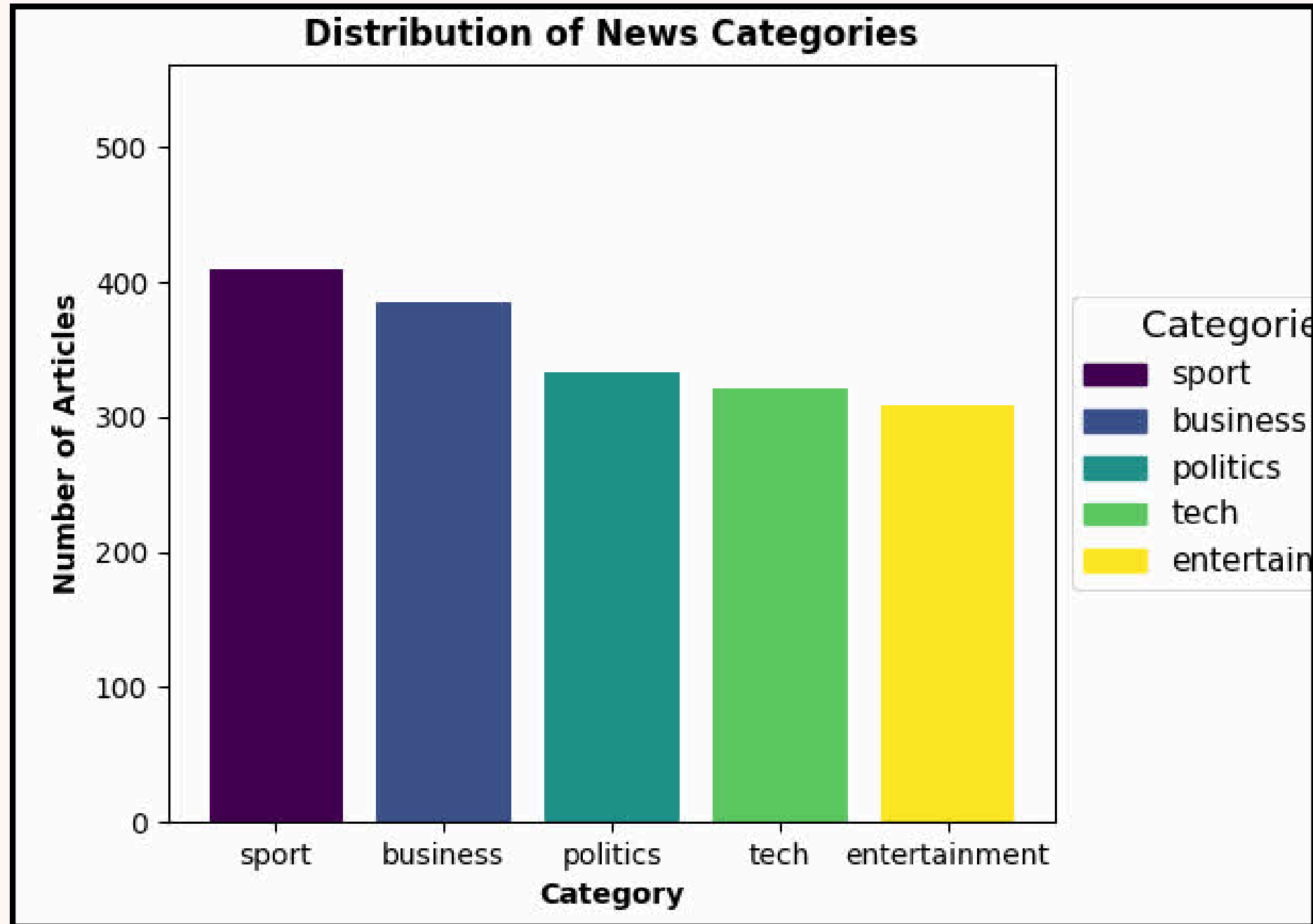
TIME CONSUMING

Long articles can be time-consuming and require a lot of concentration to get meaningful information

INFORMATION FATIGUE

Reading articles for a long period of time could lead to information fatigue, losing interest or let's just say it -
IT IS BORING

DATASET DESCRIPTION


Source: BBC dataset Kaggle

The **BBC News dataset** comprises structured directories containing **2700** text files of news articles and their respective summaries across multiple categories including business, politics, entertainment, tech, and sport.

The dataset is organized into two main folders:

- **News Articles:** Contains the full text of the news articles.
- **Summaries:** Contains the corresponding summaries for each article.

METHODOLOGY

InfoSnip utilizes advanced natural language processing techniques to extract key information from news articles and generate short summaries.

Extractive Summarization

Abstractive Summarization
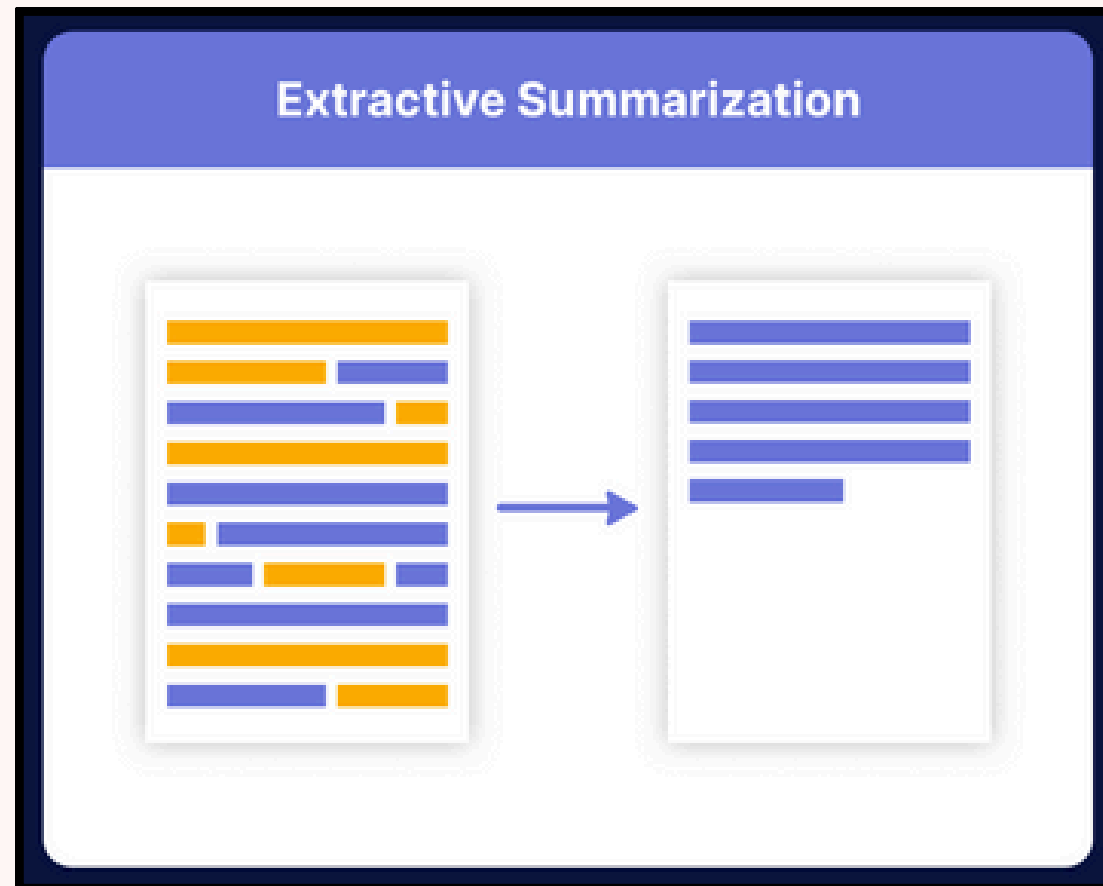
EXTRACTIVE SUMMARY

Extracts words from the news article itself to result in flash summary

TFIDF & BERT
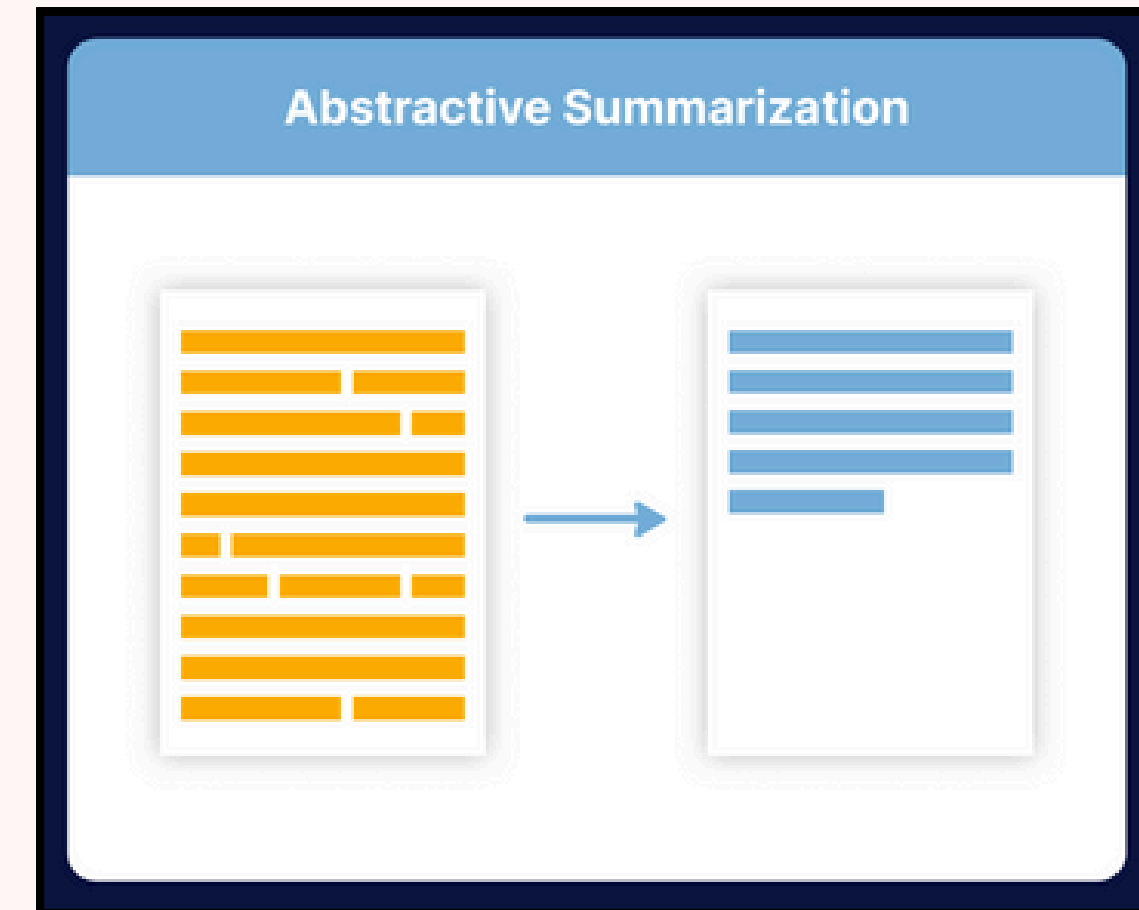
ABSTRACTIVE SUMMARY

Understands the article and provides flash summary in own words

T5 & BART

EXTRACTIVE SUMMARIZATION

BERT (**Bidirectional Encoder Representations from Transformers**)
- Encodes article sentences for deep contextual understanding
- Selects informative sentences based on semantic similarity

TF-IDF (**Term Frequency-Inverse Document Frequency**)
- Highlights word importance within documents
- Select sentences containing the highest concentration of significant terms
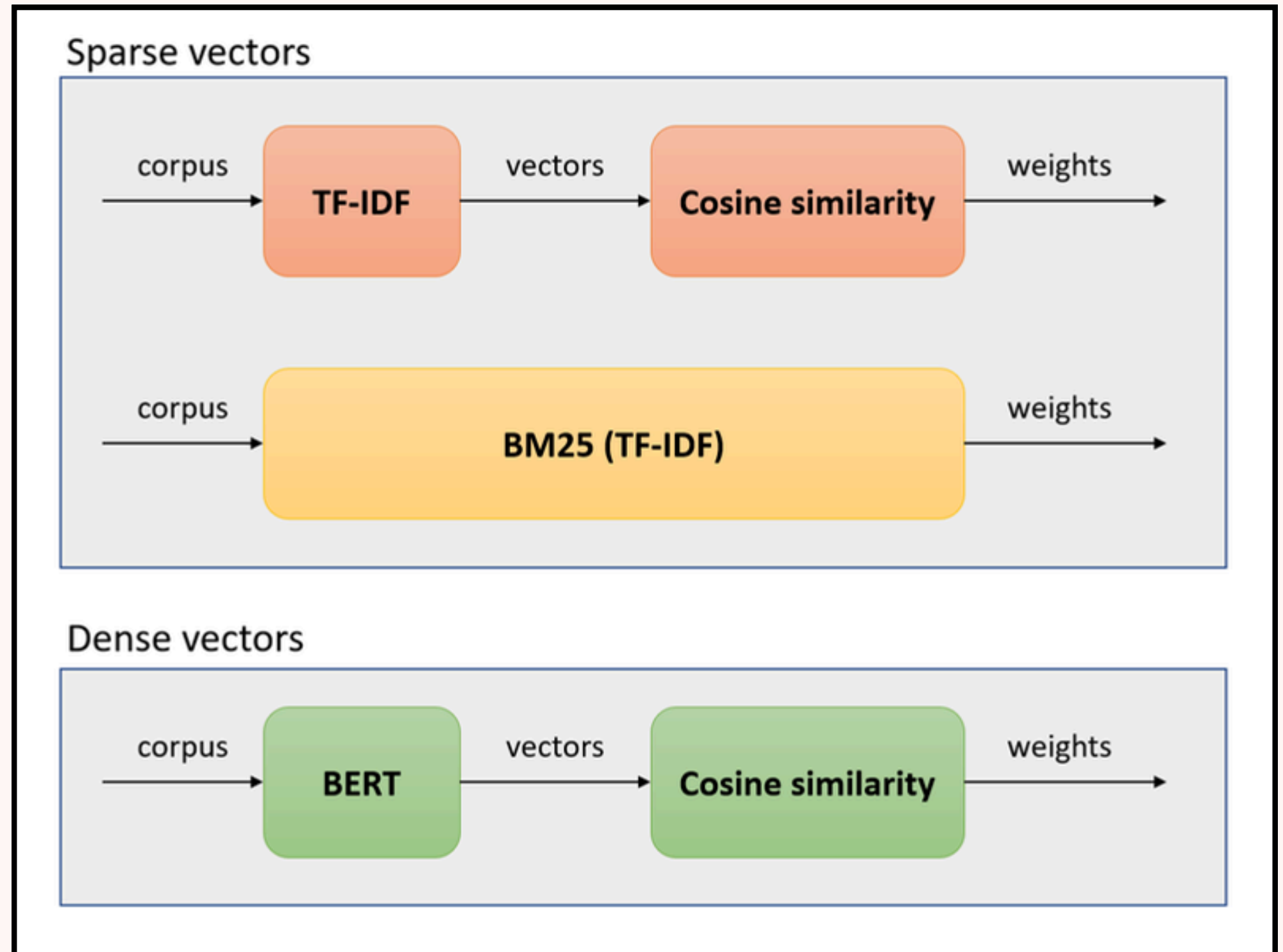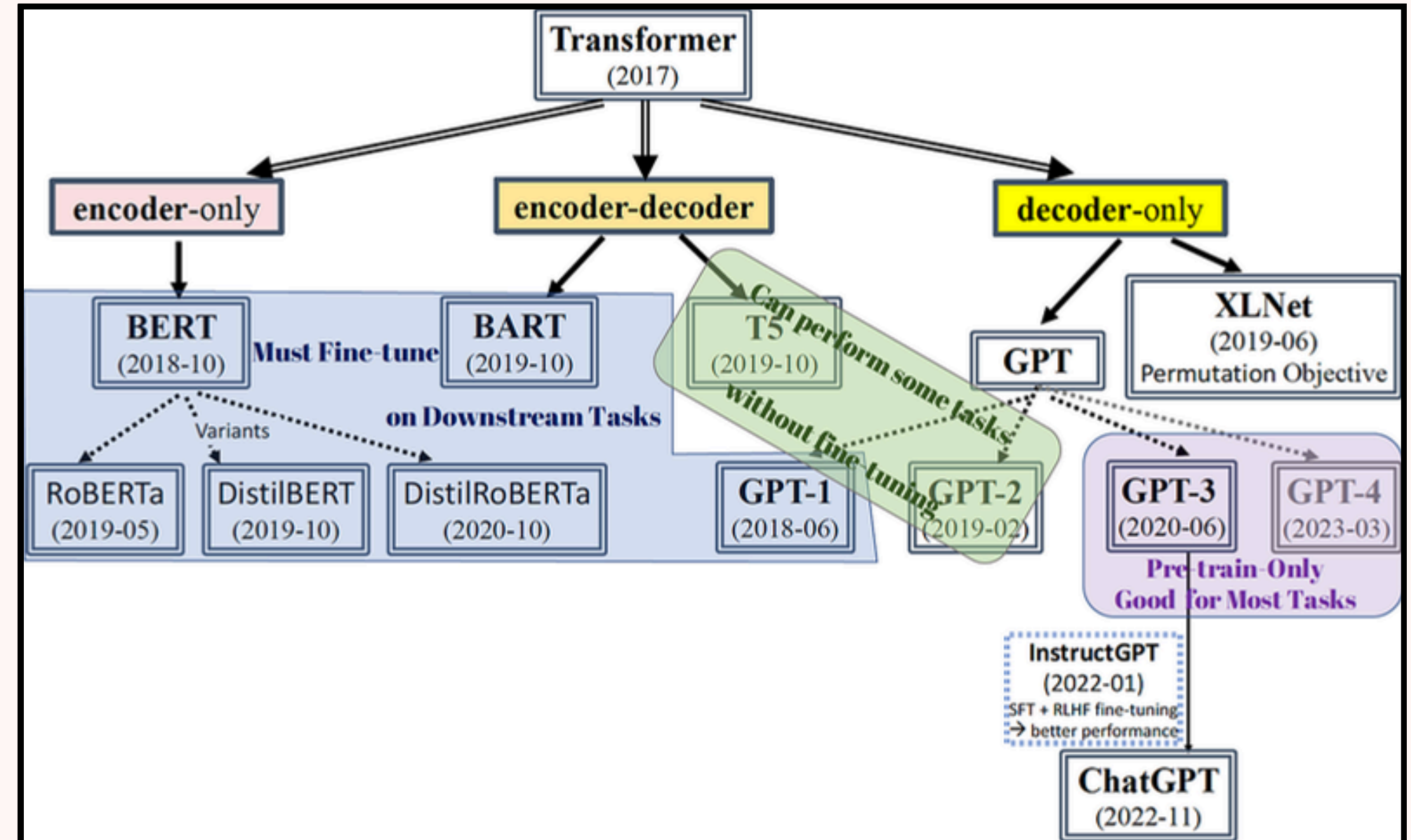
ABSTRACTIVE SUMMARIZATION

T5 (Text-to-Text Transfer Transformer)
- T5: Rewrites articles into concise form
- Paraphrases entire content, mimicking human-written abstracts

BART (Bidirectional and Auto-Regressive Transformers)
- Trained to reconstruct noisily masked text inputs
- Generates fluent, accurate summaries from lengthy or complex articles

ROUGE SCORE

- **ROUGE(Recall-Oriented Understudy for Gisting Evaluation) Scores:** Measure summary overlap with reference, e.g., ROUGE-1 for unigrams, ROUGE-2 for bigrams.
- **Precision Score:** Assess model accuracy in replicating key phrases or sentences from the original text
- **TF-IDF** demonstrates superior precision in capturing both unigrams and bigrams, as well as long-range dependencies, based on the provided scores.
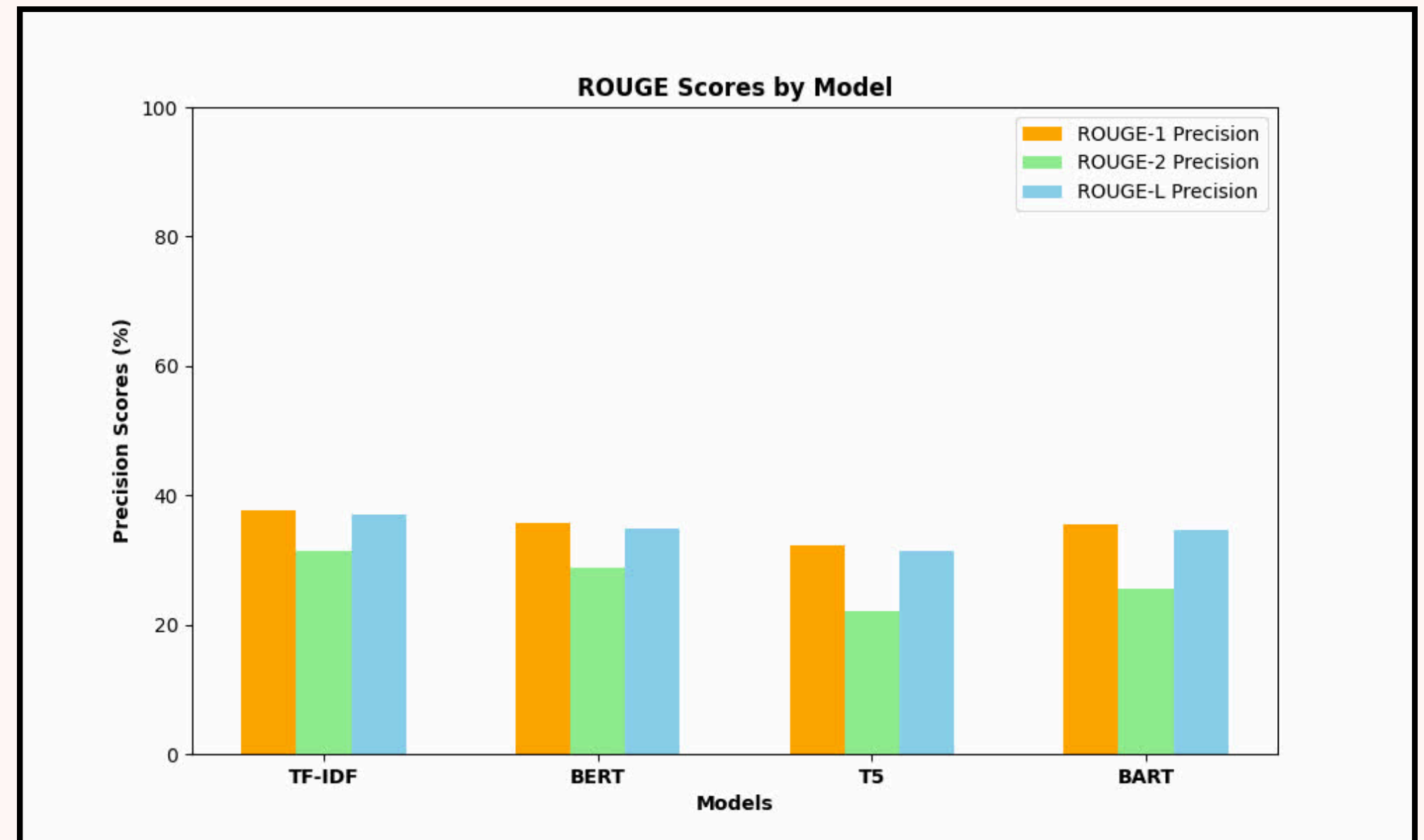
MODEL EVALUATION METRICS

- **Readability:** BART and T5 excel in producing easily understandable text, with BART slightly outperforming T5.
- **Lexical Diversity:** BART demonstrates superior performance by using a broader vocabulary, resulting in more vibrant and engaging summaries.
- **Semantic Similarity:** TF-IDF excels in this area, surpassing other models in capturing the essence and factual details of the source material.

Combined Metrics by Model
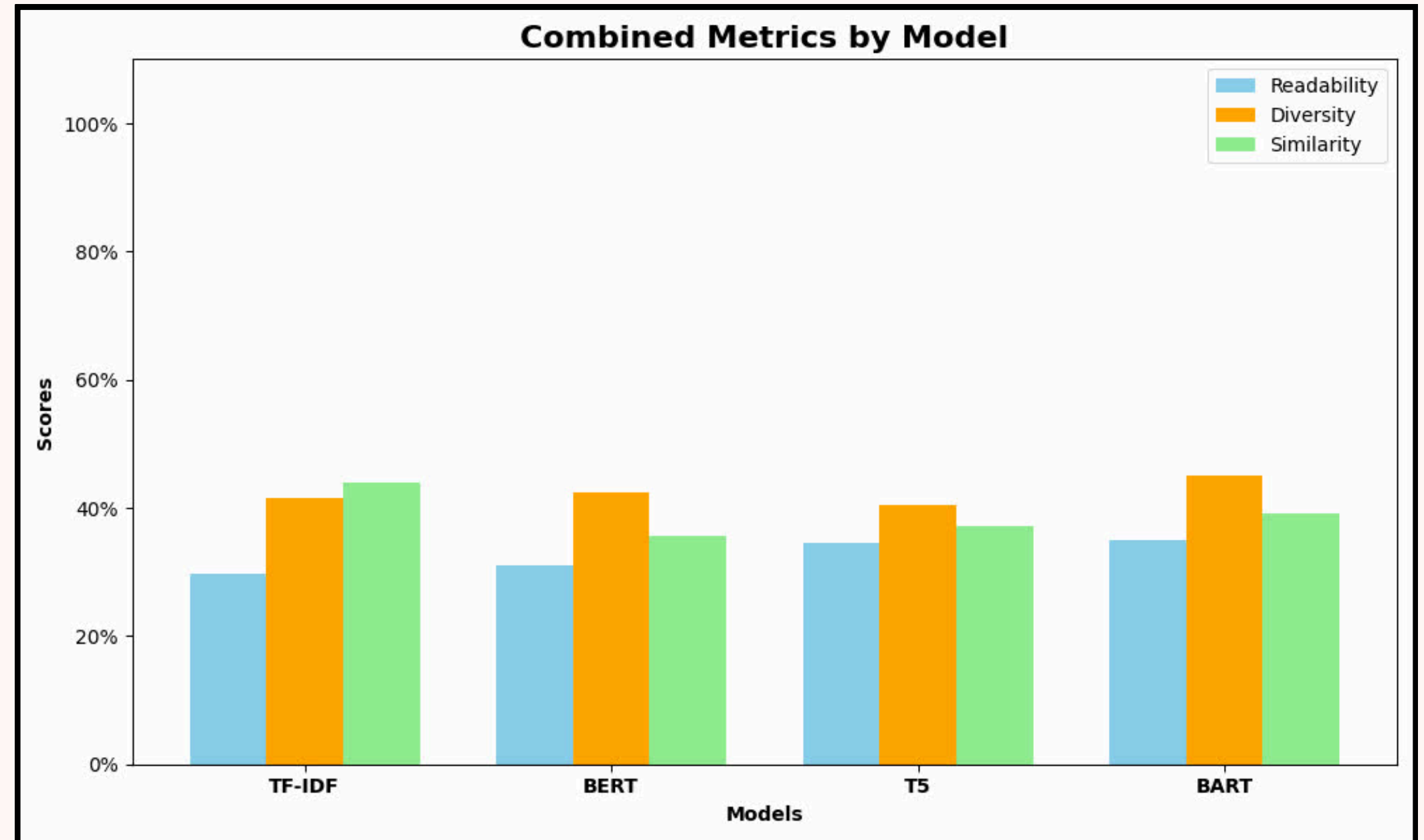
BEST MODEL?

- **BART (Bidirectional and Auto-Regressive Transformers)** delivers balanced and robust summaries, capturing essential information efficiently.
- BART's readability and varied summaries cater to diverse categories in writing concise summaries.
- **TF-IDF (Term Frequency-Inverse Document Frequency)** is preferred for maintaining original content and the long-range dependencies from the original text.

TRANSFORMERS-BART MODEL

BART

SUMMARY
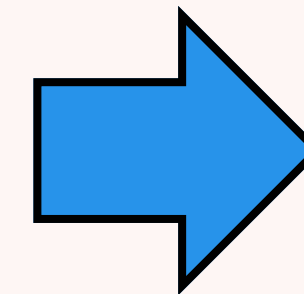
NLP
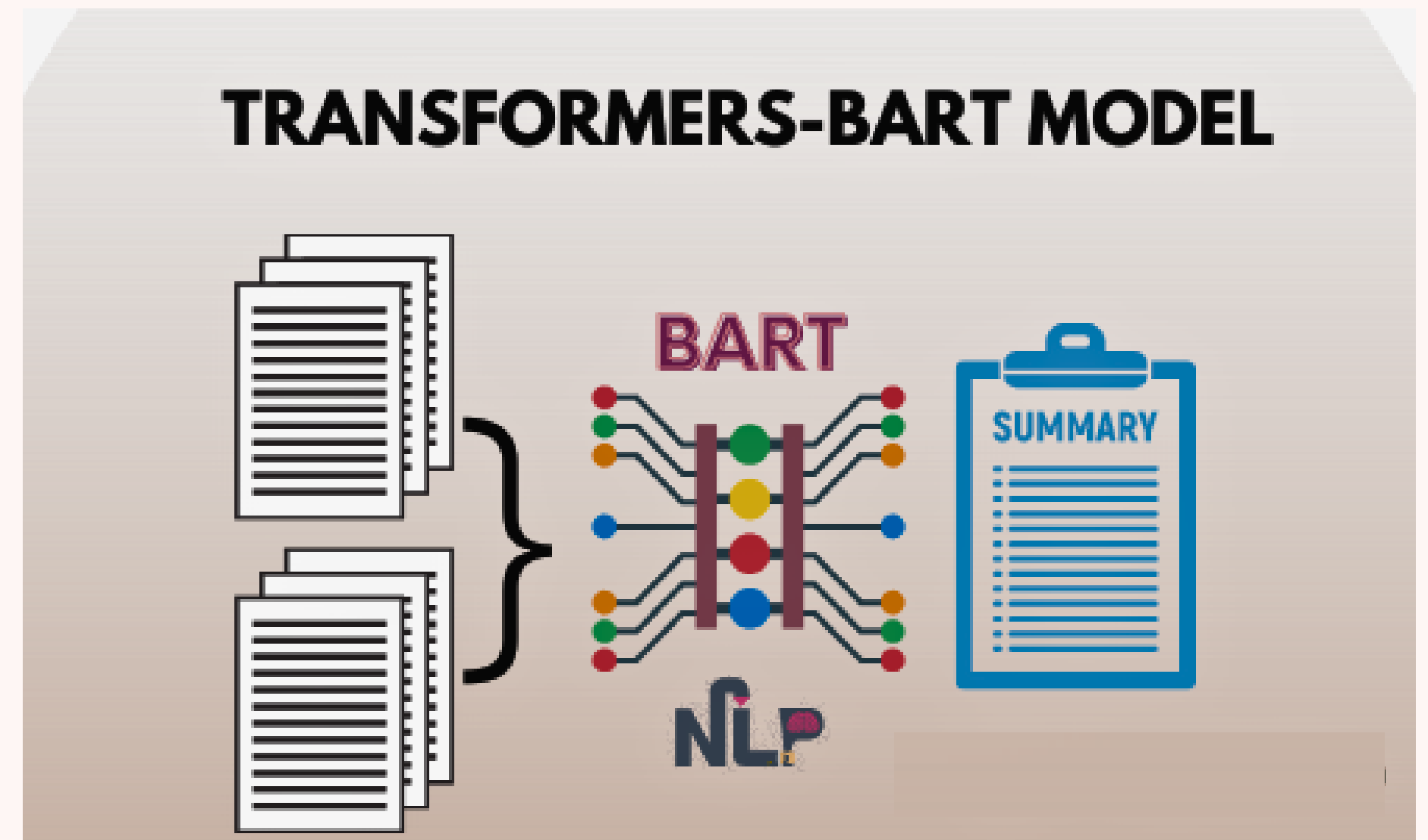
⚡ FLASH SUMMARY ⚡

InfoSnip - Tailored News Summarization

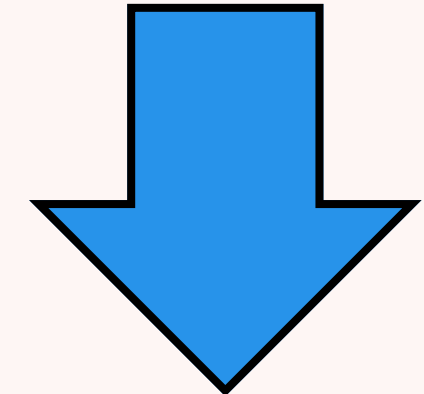**Transforming extensive articles into flash summaries**

A recent research hinted at the significant presence of methane ice within Uranus and Neptune which have long been regarded as icy giants, presumed to be predominantly composed of frozen water.
According to Space.com, the implications of this discovery extend beyond mere composition, potentially unlocking the mysteries surrounding the formation of Uranus and Neptune. Voyager 2's solitary flyby in the 1980s provided a glimpse into Uranus and Neptune, leaving scientists with only speculative notions of their makeup, including substantial amounts of oxygen, carbon, and hydrogen.

Astronomers have constructed models based on data from Voyager 2 and terrestrial telescopes to delve deeper into the composition of these distant giants. These models typically posit a thin envelope of hydrogen and helium, a dense layer of superionic water and ammonia, and a rocky core, earning the planets their "ice giant" designation. However, they also suggest the possibility of vast reservoirs of water, potentially exceeding Earth's oceans by tens of thousands of times.

The latest study challenges these conventional models by considering the formation process of Uranus and Neptune. As these planets coalesced from the primordial solar nebula, they accumulated planetesimals, reminiscent of contemporary comets from the Kuiper Belt. Unlike the presumed water-rich composition of the ice giants, these planetesimals are carbon-rich.

SUMMARIZE

Word Count Comparison

	Word Count
Original Article	388
BART Summary	32

Text Type
- Original Article
- BART Summary
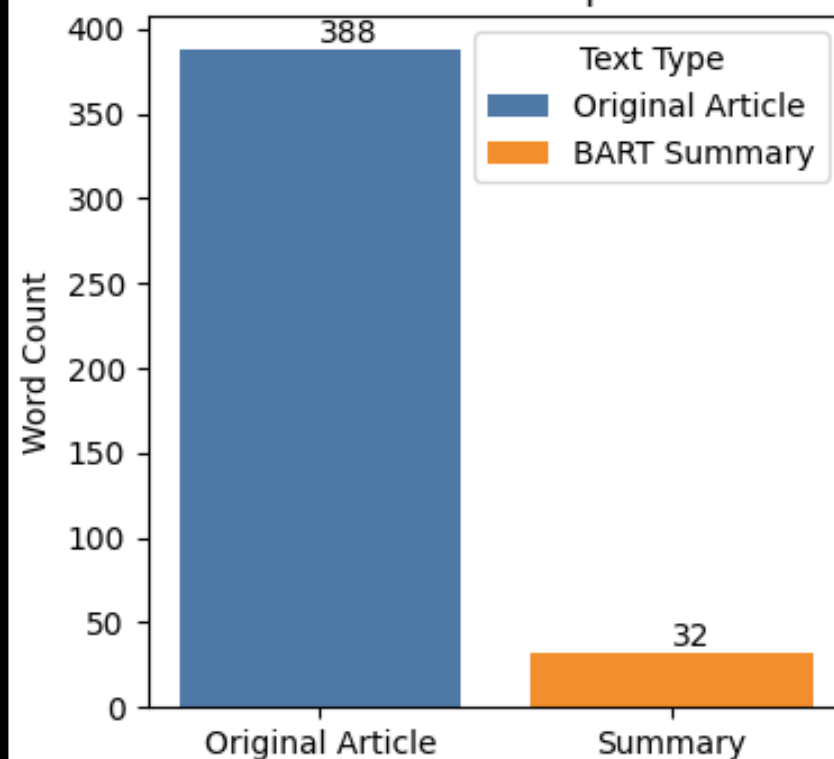
BART MODEL SUMMARY

Research suggests methane ice may have formed during the planets' tumultuous formation. Voyager 2's solitary flyby in the 1980s provided a glimpse into Uranus and Neptune, leaving scientists with only speculative notions.

SUMMARY COMPARISON WITH GPT-4

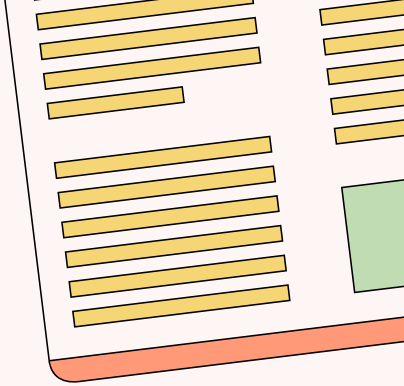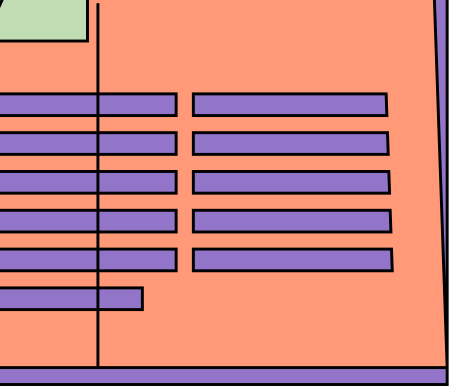| BART	CHATGPT-4
Mondo Duplantis is the reigning Olympic and world champion in the pole vault. He has broken the world record a staggering seven times in his career. The 26-year-old is the favorite to win a gold medal at the Paris Olympics in July.	**ChatGPT** Mondo Duplantis, a pole vaulting prodigy, has broken the world record seven times and holds the record at 6.23 meters. He won Olympic gold and four world titles, competes for Sweden, and aims for further records and Olympic success.
Switzerland says UBS may need more cash. The bank is fuming. The Swiss government says it may have to find as much as $27 billion to absorb potential losses. UBS chairman Colm Kelleher says he is "seriously concerned" about current discussions.	**ChatGPT** Switzerland suggests UBS may need up to $27 billion more to cover potential losses, which UBS contests, stating its financial health is strong. This regulatory proposal could impact UBS's stock and Switzerland's financial market reputation.

CONCLUSION

- Infosnip would be a one-stop solution for all the news enthusiasts.
- It gives us a condensed version of the news article which could save us a lot of time.
- InfoSnip effectively balances readability with factual accuracy

FUTURE SCOPE

- ○ **Multi-language Support**

- ○ **Audio Summaries**

- ○ **Personalized Recommendations**

- ○ **Real-Time Summarization**

THANKS FOR LISTENING!

The End