**Literature Review**

Text summarization in natural language processing (NLP) aims to produce condensed representations of extensive texts while retaining critical information. Summarization techniques are generally categorized into extractive and abstractive approaches. Extractive summarization involves selecting key sentences or phrases directly from the source text and forming a summary with original words and syntax from the input. In contrast, abstractive summarization rephrases and condenses the text, imitating human summarization. Extractive text summarization remains highly popular due to its simplicity and efficiency, especially in handling large datasets like the CNN/Daily Mail corpus (Karjule et al., 2023).

The evolution of NLP techniques has led to two prominent approaches for extractive summarization: traditional word embedding models such as Word2Vec and transformer-based models such as BERT. Traditional word embedding models leverage statistical relationships between words to create vector representations, while transformer-based models leverage attention mechanisms to capture semantic relationships within text. This review compares these two approaches' capabilities, limitations, and evaluation methods, specifically focusing on how each method impacts extractive text summarization.

**Traditional Word Embedding Techniques for Extractive Summarization**

Traditional word embedding techniques use statistical methods to convert words into continuous vector spaces, capturing semantic relationships based on word co-occurrence patterns within large corpora. Word2Vec, introduced by Mikolov et al. in 2013, is one such method that transforms words into dense vector representations, enabling the measurement of semantic similarity between words through distance calculations in vector space. Word2Vec achieves this through either a Continuous Bag of Words (CBOW) or a Skip-Gram model, which predicts a word from its surrounding context or predicts context words from a single word (Mikolov et al., 2013).

However, traditional embeddings face several inherent limitations. First, they cannot account for polysemy, the phenomenon where a word has multiple meanings. For example, in a sentence discussing “banking on a win,” Word2Vec would fail to distinguish “banking” as a betting term versus “banking” as related to financial institutions. Moreover, traditional embeddings do not consider contextual variations between words, which restricts their effectiveness in tasks that require deep semantic understanding, such as summarization. Divakar Yadav et al. explain that without contextualization, traditional models often struggle to capture the nuanced meanings necessary for accurate summarization, resulting in summaries that may lack coherence and relevance when applied to complex datasets (Yadav et al., 2021).

**Transformer-Based Models for Extractive Summarization**

In recent years, transformer-based models have revolutionized NLP by providing contextual word representations that significantly improve the performance of various NLP tasks, including text summarization. Transformers, introduced by Ashish Vaswani et al. in their influential paper “Attention Is All You Need,” rely on self-attention mechanisms, which allow the model to weigh the significance of each word relative to all other words in a sentence. This structure enables transformers to handle long-range dependencies and capture complex semantic relationships, making them well-suited for tasks that require a deep understanding of context (Vaswani et al., 2017).

BERT (Bidirectional Encoder Representations from Transformers), developed by Devlin et al. in 2018, builds on the transformer architecture by training a language model to consider the bidirectional context of each word. BERT captures relationships both to the left and right of each word, providing a more nuanced understanding of word meaning. This bidirectional training enables BERT to understand subtle context differences within a document, leading to more coherent and relevant summaries. Abdel-Salam and Rafea demonstrate that BERT-based models significantly outperform traditional embeddings in extractive summarization, achieving higher scores on relevance, coherence, and grammatical accuracy. This improvement is particularly evident when evaluated with metrics like ROUGE and BLEU, where BERT-based models often yield scores far surpassing those of Word2Vec-based models (Abdel-Salam & Rafea, 2022).

Despite the performance advantages, transformer models come with substantial computational costs. BERT, for instance, requires considerable memory and processing power due to its multiple attention layers and bidirectional processing. Rajesh et al. emphasize that while BERT’s complexity enables more accurate summarization, the computational trade-offs may limit its application in low-resource environments or applications requiring real-time processing. Nevertheless, techniques like distillation and pruning are being explored to reduce the computational footprint of transformer models without significant performance losses, potentially making them more accessible for a wider range of applications (Rajesh et al., 2024).

**Evaluation Metrics for Summarization Techniques**

Assessing the effectiveness of text summarization models requires robust evaluation metrics that capture aspects such as relevance, coherence, and fluency. The ROUGE (Recall-Oriented Understudy for Gisting Evaluation) metric is one of the most widely adopted methods for evaluating extractive summarization models. ROUGE calculates the overlap between n-grams in model-generated summaries and reference summaries, focusing on metrics such as precision, recall, and F1-score. Auriemma Citarella et al. explain that ROUGE is highly effective for extractive summarization, as it directly measures the extent to which generated summaries reflect the content of the original text (Auriemma Citarella et al., 2023).

Evaluation metrics can also extend beyond ROUGE, including human evaluation for coherence and informativeness. For instance, Abdel-Salam and Rafea supplemented ROUGE scores with human judgments, finding that BERT-based summaries were rated as more coherent and fluent than those generated by Word2Vec, despite the challenges of increased computational requirements (Abdel-Salam & Rafea, 2022).

**Comparative Performance Analysis and Trade-Offs**

The trade-off between model accuracy and computational efficiency is a critical consideration in selecting a text summarization approach. Traditional word embeddings like Word2Vec offer low computational requirements and faster training times, making them more feasible for applications where computational resources are limited. According to Rajesh et al., traditional embeddings are ideal for simple summarization tasks that do not require advanced contextual understanding, though their limitations become apparent with more complex datasets (Rajesh et al., 2024).

In contrast, transformer-based models offer greater performance in terms of summarization quality and relevance but require significant computational resources due to their multi-layered architecture. Abdel-Salam and Rafea report that while BERT models perform exceptionally well in producing contextually accurate summaries, their training times and memory demands are considerably higher than those of traditional models. This trade-off presents a dilemma for researchers and practitioners who must balance the need for high-quality summarization with computational feasibility, particularly in environments with limited resources (Abdel-Salam & Rafea, 2022).

Ongoing research seeks to address these challenges by optimizing transformer models for efficiency. Techniques such as distillation—reducing model size while retaining performance—have shown promise in making transformers more accessible without sacrificing much accuracy. Hybrid approaches, which combine traditional embeddings for initial stages with transformers for refinement, also present potential solutions, offering a compromise between efficiency and accuracy (Karjule et al., 2023).

**Conclusion**

The literature reveals distinct advantages and limitations of traditional and transformer-based models for extractive summarization. Traditional models like Word2Vec provide a simpler, computationally efficient approach, suitable for applications where resources are limited. However, their lack of contextual understanding often limits the quality of summaries. Transformer-based models like BERT, while computationally intensive, offer superior performance in generating coherent, contextually accurate summaries, proving especially effective on complex datasets. The choice of model thus depends on the specific requirements of the application, with transformer models excelling where accuracy and context are paramount and traditional models serving as a feasible alternative where computational efficiency is essential.