

# Research on DistilBERT and ALBERT

Abdelrahman Ahmed

September 19, 2025

## 1 Introduction

BERT (Bidirectional Encoder Representations from Transformers) brought major improvements to NLP tasks but is large and computationally expensive. DistilBERT and ALBERT were developed to address these issues while maintaining high performance.

## 2 DistilBERT

**Definition:** DistilBERT is a compressed version of BERT created using Knowledge Distillation. It is 40% smaller, 60% faster, and retains about 97% of BERT's performance.

**Knowledge Distillation:** DistilBERT is trained by imitating a large BERT teacher model, learning not only from labels but also from the probability distributions of the teacher.

**Applications:** Text classification, question answering, and named entity recognition.

## 3 ALBERT

**Definition:** ALBERT (A Lite BERT) drastically reduces the number of parameters through techniques like Weight Sharing and Factorized Embedding Parameterization.

**Features:** Smaller embedding matrices, shared parameters across layers, and efficient training. ALBERT often achieves equal or better performance than BERT.

**Applications:** General NLP tasks, especially in resource-constrained environments.

## 4 Comparison

| Feature            | DistilBERT             | ALBERT                                  |
|--------------------|------------------------|---|
| Compression Method | Knowledge Distillation | Weight Sharing + Factorized Embedding   |
| Model Size         | 40% smaller than BERT  | Much smaller (due to parameter sharing) |
| Speed              | 60% faster than BERT   | Faster training & inference             |
| Accuracy           | ~97% of BERT           | Sometimes better than BERT              |
| Best Use Case      | Prioritize speed       | Prioritize memory efficiency            |

## 5 Conclusion

DistilBERT offers near-BERT accuracy with higher speed, while ALBERT is optimal when memory efficiency is critical. Both models demonstrate how to scale down BERT for practical use cases.