Web Summarizer Model Testing Results

Comprehensive Model Performance Analysis

Generated on: June 21, 2025 at 02:50 AM

# Executive Summary

This document presents comprehensive testing results for the Web Summarizer application. Four different AI models were tested across various text lengths to evaluate their performance in terms of speed, accuracy, and compression ratio. All models achieved 100% success rate, demonstrating the reliability of the summarization system.

## Table 1: Overall Performance Summary

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **Success Rate** | **Avg Time (s)** | **Avg Similarity** | **Avg Compression** |
| BART | 100.0% | 8.442s | 95.1% | 58.3% |
| T5 | 100.0% | 5.886s | 79.6% | 48.3% |
| PEGASUS | 100.0% | 15.808s | 72.8% | 56.9% |
| Arabic (mT5) | 100.0% | 10.216s | 71.4% | 36.1% |

## Table 2: Detailed Results by Text Length

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Text Length** | **Model** | **Success** | **Time (s)** | **Similarity** | **Compression** | **Words** |
| Short | BART | ✓ | 12.771s | 98.8% | 127.6% | 51 |
| Short | T5 | ✓ | 6.051s | 89.4% | 105.8% | 44 |
| Short | PEGASUS | ✓ | 11.037s | 83.7% | 121.0% | 46 |
| Short | Arabic | ✓ | 11.376s | 68.2% | 78.2% | 33 |
| Medium | BART | ✓ | 6.326s | 91.7% | 35.6% | 47 |
| Medium | T5 | ✓ | 4.309s | 68.8% | 27.6% | 45 |
| Medium | PEGASUS | ✓ | 10.875s | 53.9% | 28.6% | 37 |
| Medium | Arabic | ✓ | 8.714s | 66.4% | 22.6% | 32 |
| Long | BART | ✓ | 6.228s | 94.9% | 11.7% | 46 |
| Long | T5 | ✓ | 7.299s | 80.6% | 11.5% | 46 |
| Long | PEGASUS | ✓ | 25.511s | 80.8% | 21.2% | 80 |
| Long | Arabic | ✓ | 10.559s | 79.6% | 7.5% | 26 |

## Table 3: Model Rankings

|  |  |  |  |
| --- | --- | --- | --- |
| **Category** | **Best Model** | **Score** | **Details** |
| 🏃 Speed | T5 | 5.886s | Fastest processing time |
| 🎯 Accuracy | BART | 95.1% | Highest semantic similarity |
| 📝 Compression | Arabic | 36.1% | Best text compression |
| 🏆 Overall | BART | 16.262 | Best balanced performance |

## Table 4: Model Specifications

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **Architecture** | **Parameters** | **Training Data** | **Specialization** |
| BART | Encoder-Decoder | 400M | CNN/DailyMail | News summarization |
| T5 | Encoder-Decoder | 220M | C4 corpus | General text tasks |
| PEGASUS | Encoder-Decoder | 568M | C4 + HugeNews | Abstractive summarization |
| Arabic (mT5) | Encoder-Decoder | 580M | Multilingual | Arabic text processing |

## Table 5: Key Findings

|  |  |  |  |
| --- | --- | --- | --- |
| **Finding** | **Model** | **Value** | **Significance** |
| Fastest Processing | T5 | 5.886s | Best for real-time applications |
| Most Accurate | BART | 95.1% | Best for quality-critical tasks |
| Best Compression | Arabic | 36.1% | Most concise summaries |
| Most Reliable | All | 100% | All models work consistently |
| Best for English | BART | 95.1% | Excellent accuracy |
| Best for Arabic | mT5 | 71.4% | Specialized for Arabic text |

## Table 6: Recommendations by Use Case

|  |  |  |
| --- | --- | --- |
| **Use Case** | **Recommended Model** | **Reason** |
| Real-time Applications | T5 | Fastest processing |
| High-Quality Summaries | BART | Highest accuracy |
| Arabic Content | mT5 | Specialized for Arabic |
| General Purpose | BART | Balanced performance |
| News Articles | BART | Trained on news data |
| Academic Papers | PEGASUS | Better for complex texts |

## Testing Methodology

The testing was conducted using three different text lengths:  
  
**•** Short texts (150-300 words): News articles and brief descriptions  
**•** Medium texts (800-1200 words): Detailed articles and reports  
**•** Long texts (2000+ words): Comprehensive documents and research papers  
  
Metrics measured:  
  
**•** Processing Time: Time from input to summary generation (seconds)  
**•** Semantic Similarity: How well the summary preserves original meaning (0-100%)  
**•** Compression Ratio: How much text is condensed (percentage of original)  
**•** Success Rate: Percentage of successful summarizations (0-100%)

## Conclusions

Based on the comprehensive testing results:  
  
**1.** All models achieved 100% success rate, demonstrating high reliability  
  
**2.** BART emerges as the best overall model with 95.1% accuracy  
  
**3.** T5 is the fastest model, making it ideal for real-time applications  
  
**4.** Arabic (mT5) provides specialized support for Arabic text processing  
  
**5.** The system successfully handles multiple languages and text lengths  
  
**6.** Performance varies by text length, with models showing different strengths