Ranking summary evaluation techniques can be somewhat subjective and dependent on the specific use case and requirements of the evaluation. However, based on general performance and adoption in the field of Natural Language Processing (NLP), here's a ranked list with reasons:

1. **GPTSCORE**
2. **BERTScore**
3. **UniEVAL**
4. **QuestEval**
5. **ROUGE L**
6. **METEOR**
7. **ROUGE 2**
8. **ROUGE 1**
9. **ROUGE S**
10. **BLEU**

### **Reasons for Ranking:**

1. **GPTSCORE**:
   * **Best**: Utilizes advanced language models like GPT to provide comprehensive evaluation by understanding the context and semantics of the text. It captures deeper linguistic features and provides more human-like evaluations.
2. **BERTScore**:
   * **Very Good**: Uses BERT embeddings to evaluate summaries based on contextual word embeddings, capturing semantic similarity better than traditional metrics.
3. **UniEVAL**:
   * **Very Good**: A newer evaluation method that integrates various aspects of evaluation (e.g., content, fluency) into a unified framework, providing a balanced and comprehensive assessment.
4. **QuestEval**:
   * **Good**: Focuses on question-answering to evaluate the quality of the summary, ensuring that key information is retained and relevant questions can still be answered from the summary.
5. **ROUGE L**:
   * **Good**: Considers longest common subsequences, which is beneficial for capturing fluency and sequence of the summary.
6. **METEOR**:
   * **Average**: Focuses on precision and recall but also includes synonymy and stemming, making it better than purely n-gram based methods.
7. **ROUGE 2**:
   * **Average**: Measures bigram overlap, capturing some level of fluency and coherence but less effective in capturing semantic similarity.
8. **ROUGE 1**:
   * **Below Average**: Measures unigram overlap, providing a basic measure of content overlap but missing context and deeper semantic understanding.
9. **ROUGE S**:
   * **Below Average**: Measures skip-bigram overlap, which can be useful but is generally less intuitive and harder to interpret compared to other ROUGE metrics.
10. **BLEU**:
    * **Worst**: Originally designed for machine translation, focuses heavily on precision, and can penalize creative or valid variations in wording, making it less suitable for summary evaluation.

### **Tabular Representation:**

| **Rank** | **Evaluation Technique** | **Reasoning** |
| --- | --- | --- |
| 1 | GPTSCORE | Best for capturing deep linguistic features and providing human-like evaluations. |
| 2 | BERTScore | Excellent for semantic similarity using contextual embeddings. |
| 3 | UniEVAL | Comprehensive evaluation integrating multiple aspects. |
| 4 | QuestEval | Ensures key information retention through question-answering. |
| 5 | ROUGE L | Captures fluency and sequence through longest common subsequences. |
| 6 | METEOR | Balances precision, recall, synonymy, and stemming. |
| 7 | ROUGE 2 | Measures bigram overlap, useful but limited in semantic capture. |
| 8 | ROUGE 1 | Basic content overlap measure, lacks context understanding. |
| 9 | ROUGE S | Skip-bigram overlap, less intuitive and harder to interpret. |
| 10 | BLEU | Designed for translation, overly focused on precision, not ideal for summaries. |

This ranking aims to provide a general perspective, but the best evaluation method might vary depending on specific project needs and the nature of the text being evaluated.