# Process for Evaluation

---

## Evaluating Semantic Relationship

### Bertscorer

\* Compares word embeddings using BERT,  
\* Measures similarity by aligning words in the generated summary to the source material, takes context into account.  
\* Rewards semantically similar words (synonyms) and penalizes mismatches.  
\* More sensitive to word meaning and placement.

### Cosign Similarity

\* Compares sentence level embeddings, vector representation of the whole sentence.  
\* Checks the overall semantic meaning, not just for word for word matching like BertScorer.  
\* More forgiving to text variations as long as the core meaning is the same.  
  
Formula:  
 Precision = LCS length / number of words in the generated summary  
 Recall = LCS length / number of words in the reference summary  
 F1 Score combines both to balance precision and recall.

---

## Evaluating Lexical Overlap

### Rouge\_L

\* Measures longest common subsequence of words (LCS) between the generated summary and source documents.  
\* Focuses on word order, the longer the sequence of words in the correct order the higher the score.  
  
\* Precision: Number of words in LCS / Number of words in the generated output  
\* Recall: Number of words in LCS / Number of words in the reference summary

---

## Evaluating Relavance

### Named based Entities (Spacy)

\* Extracts named entities (e.g., people, organizations, locations, dates, bill names).  
\* Measures if key factual elements from source document are in the generated summary, ensuring factual accuracy.