**Name: Suhaila Ahmed Hassan**
**Track: AI**
**Branch: Alexandria**

# N-ROUGE

ROUGE stands for Recall-Oriented Understudy for Gisting Evaluation and the N part denotes N-gram.
It is a metric used to evaluate the quality of text generated by a machine learning model, particularly in tasks such as text summarization and machine translation.
It measures the n-gram overlap between the generated text and reference text

## Key Concepts:

**N-grams:**
In ROUGE-N, "N" refers to the length of the sequence of words that are being compared between the candidate and reference text.
For example:

- ROUGE-1 compares unigrams (single word).
- ROUGE-2 compares bigrams (two consecutive words).
- ROUGE-3 compares trigrams (three consecutive words), and so on.

**Recall:**
ROUGE-N primarily focuses on recall, which means how much the words in the human references appear in the candidate model outputs.

## Example:

Human-Produced Reference: Dan loves chocolate cakes
Model-Generated Candidate: Dan loves chocolate chip cookies
ROUGE-1:
ROUGE[1]-recall = 3/4 = 0.75
ROUGE[1]-precision = 3/5 = 0.6
ROUGE[n]-F1 = 2*(0.75*0.6)/(0.75+0.6) = 0.66

Example Source: traceloop.com

## Advantages:

- Well-established metric.
- Easy to calculate
- Easy to understand.
- Can be used to evaluate summaries in any language.

## Disadvantages:

- Only measures n-gram overlap and does not take into account the semantic meaning of the summary.
- Sensitive to the choice of reference summaries.
- Can be biased towards summaries that are shorter or longer than the reference summaries.

**Resources**

https://medium.com/@eren9677/text-summarization-387836c9e178
https://www.traceloop.com/blog/evaluating-model-performance-with-the-rouge-metric-a-comprehensive-guide