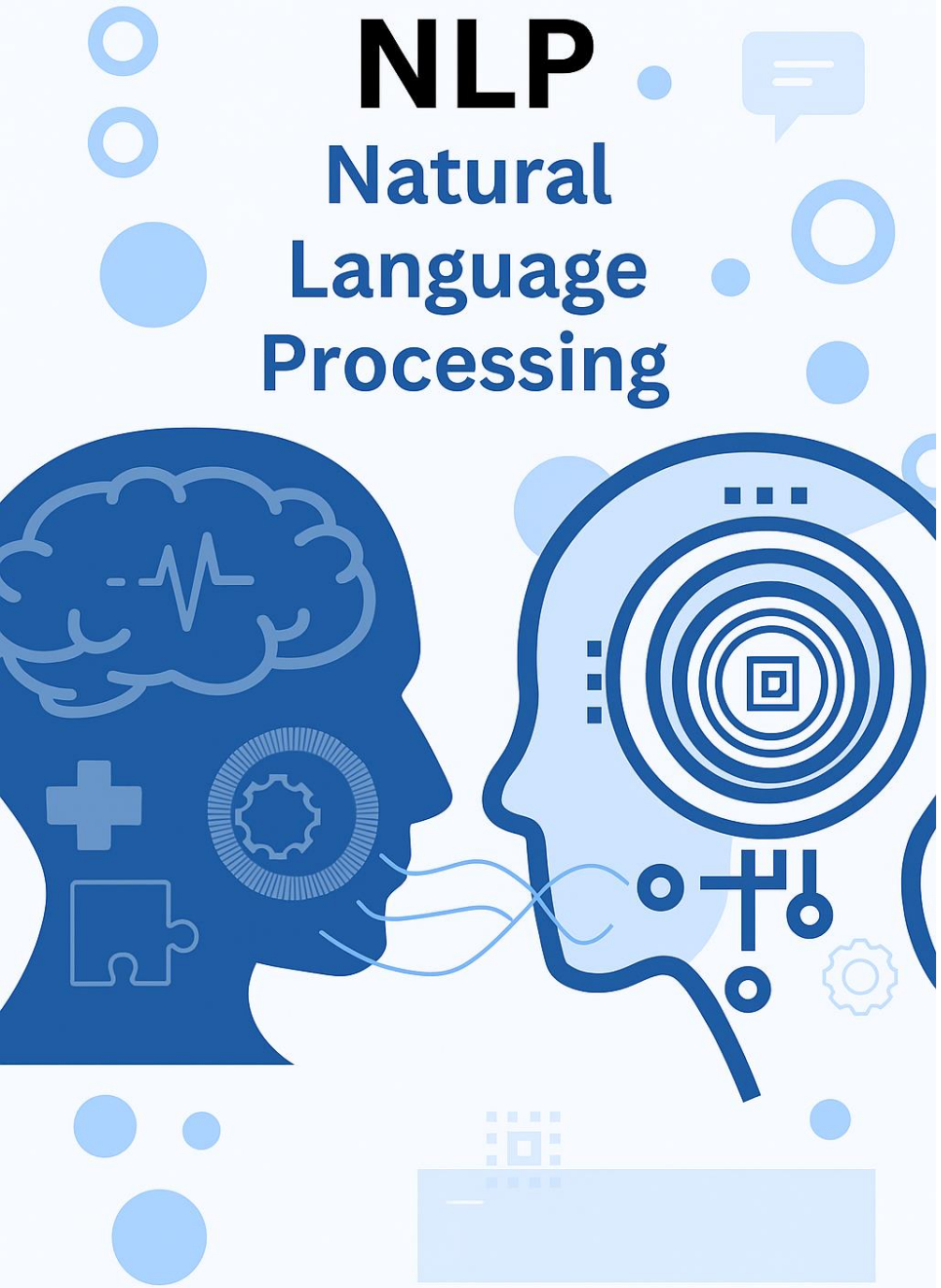

NATURAL LANGUAGE PROCESSING (NLP)

PMDS606L

MODULE 2

LECTURE 2

Dr. Kamanasish Bhattacharjee

Assistant Professor

Dept. of Analytics, SCOPE, VIT

MINIMUM EDIT DISTANCE

- Minimum Edit Distance between two strings is the **minimum number of edit operations** (insertions, deletions, substitutions) needed to convert one string into another.

MINIMUM EDIT DISTANCE

- **Levenshtein Distance**: All three operations allowed.
- **Levenshtein Distance without Substitution**: Insertion, Deletion allowed. Substitution not allowed.
- **Damerau-Levenshtein Distance**: Also allows **transpositions** (swapping two adjacent characters).
- **Hamming Distance**: Only substitutions, and strings must be of equal length.

Hamming Distance = $||A \oplus B||$ = 3

WAGNER-FISCHER ALGORITHM

- https://medium.com/@yasaswini.gaddam21/wagner-fischer-algorithm-minimum-edit-distance-4e61bba9b656
- https://www.youtube.com/watch?v=We3YDTzNXEk
- https://calc.hypotheses.org/3265