



Natural Language Processing with Deep Learning

Bài 1: GIỚI THIỆU NLP



https://csc.edu.vn/data-science-machine-learning/natural-language-processing-with-deep-learning_293

2023



GIỚI THIỆU NLP



I. Tổng quan Natural Language Processing

II. Chức năng của NLP

III. Natural Language Processing Levels

IV. Ứng dụng NLP



Lịch sử Natural Language Processing

- Bắt đầu phát triển từ những năm 1950.
- Phát triển từ các phương pháp cơ bản như xử lý từ vựng và ngũ pháp đến các công nghệ tiên tiến như các **mô hình ngôn ngữ** và **Deep Learning**.
- Đóng vai trò quan trọng trong quá trình xử lý và hiểu ngôn ngữ tự nhiên.





Lịch sử Natural Language Processing

NATURAL LANGUAGE PROCESSING

Là ngôn ngữ phát triển tự nhiên thông qua quá trình sử dụng của con người

Ví dụ: tiếng Việt, tiếng Anh, v.v.

KHÔNG PHẢI là ngôn ngữ nhân tạo (*Klingon*) hay ngôn ngữ lập trình (*Python*).



GIỚI THIỆU NLP



I. Tổng quan Natural Language Processing

II. Chức năng của NLP

III. Natural Language Processing Levels

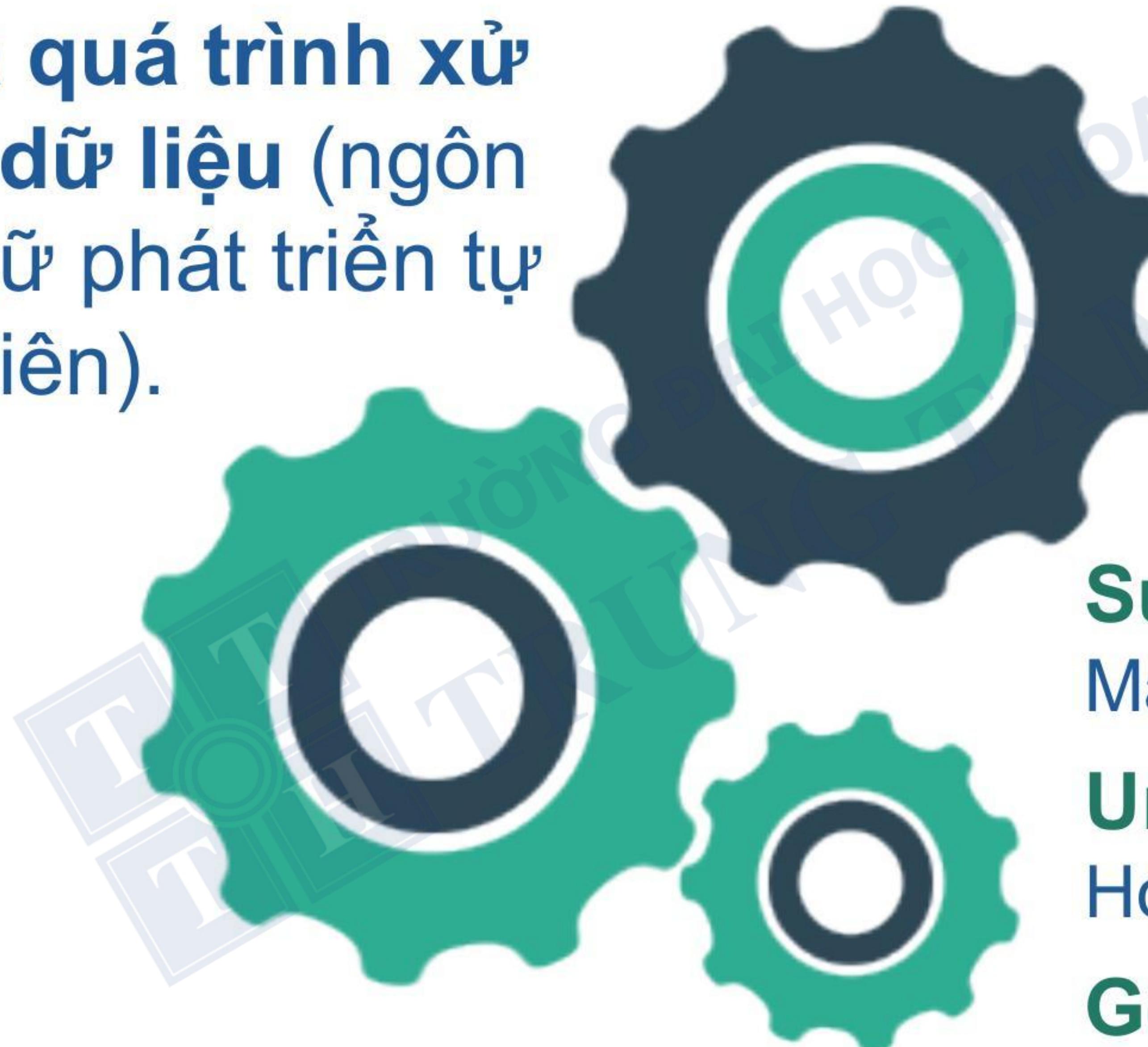
IV. Cấu trúc NLP



Chức năng của NLP

NATURAL LANGUAGE PROCESSING

Là quá trình xử lý dữ liệu (ngôn ngữ phát triển tự nhiên).

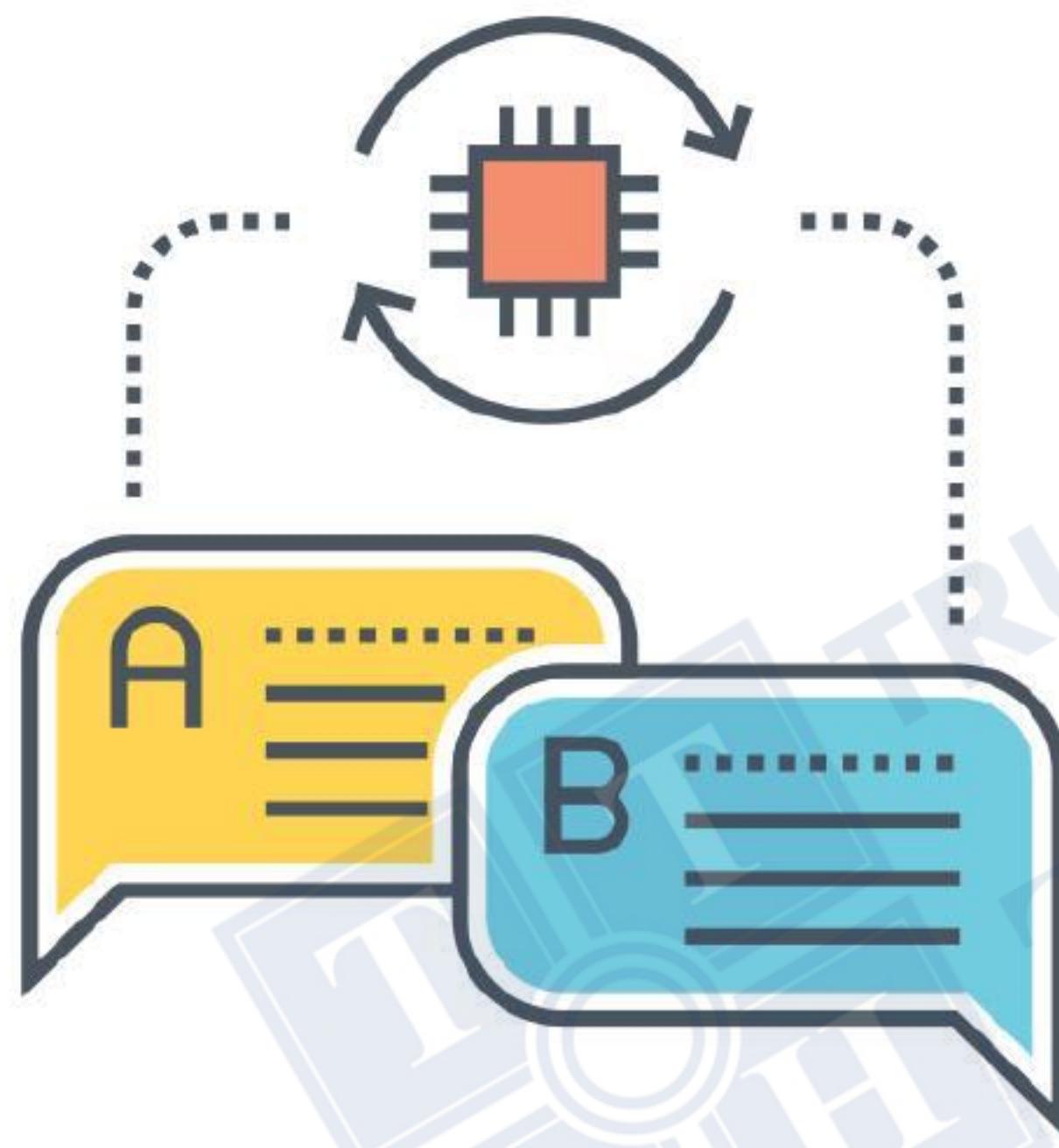


- Supervised learning**
Map văn bản với X
- Unsupervised learning**
Học X từ văn bản
- Generate** văn bản từ X



Chức năng của NLP

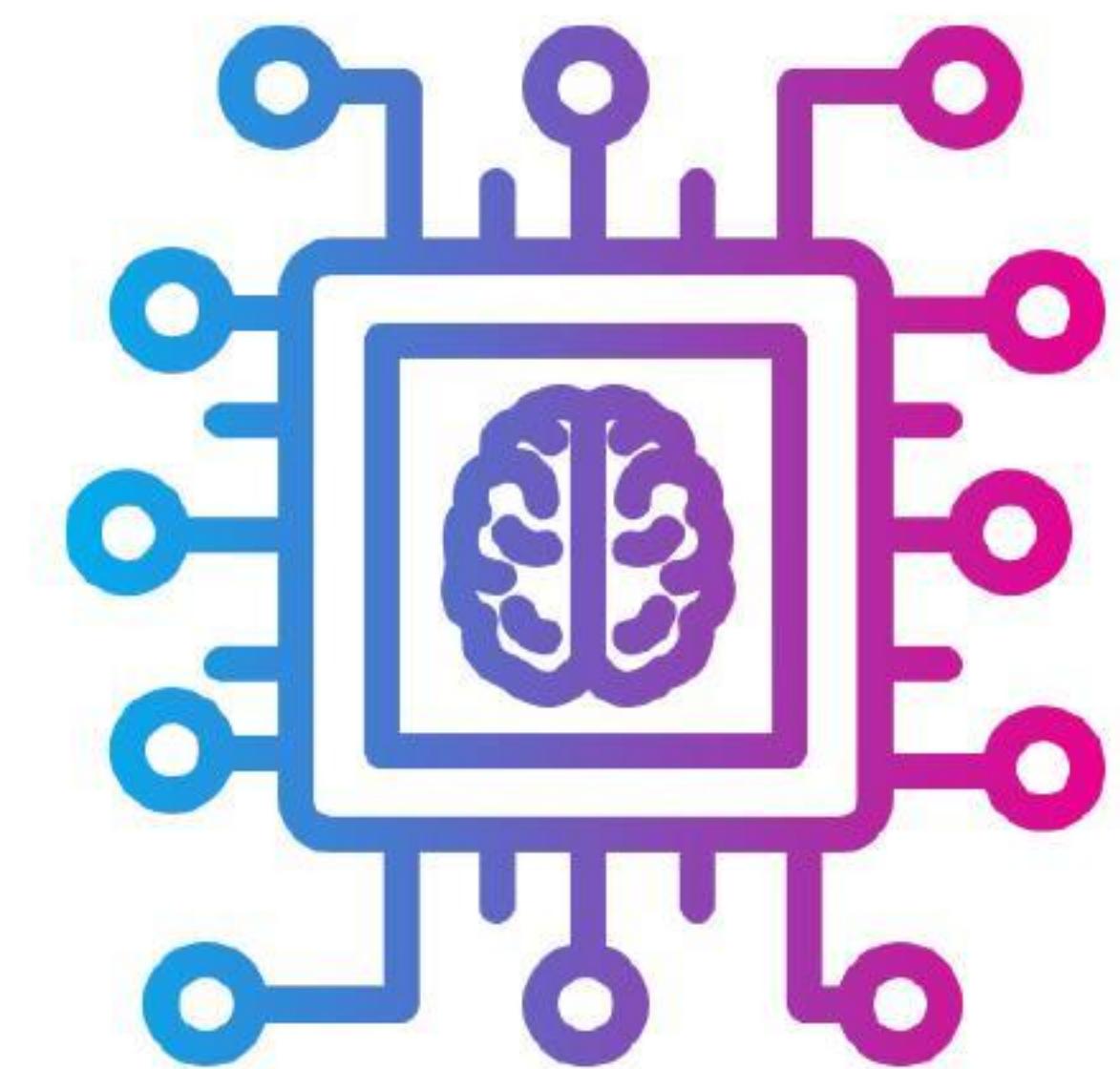
- Natural Language Processing (NLP) là quá trình xử lý và hiểu ngôn ngữ tự nhiên bằng cách ứng dụng trí tuệ nhân tạo. Bao gồm các khái niệm:



Natural Language Processing

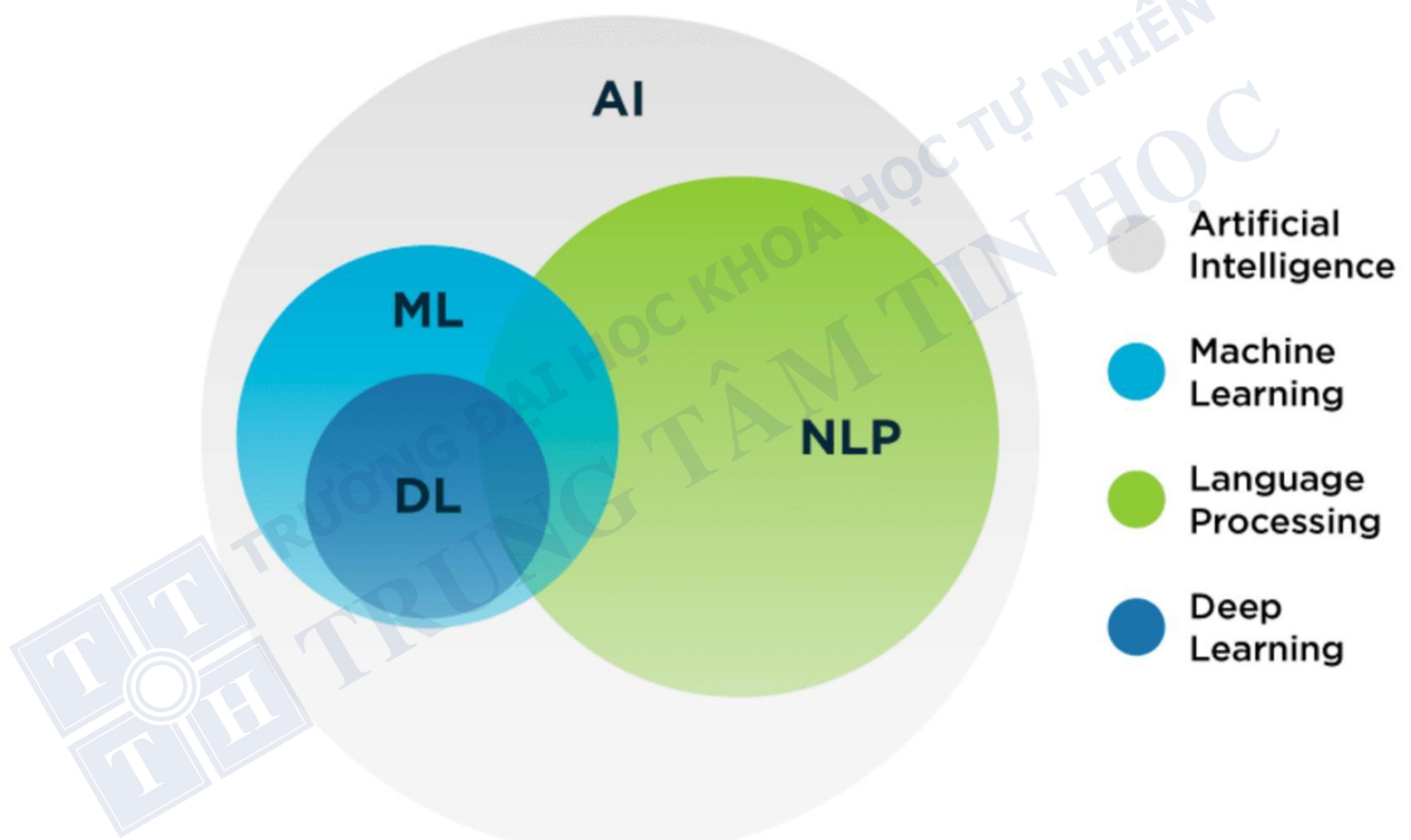


Computational Linguistics



Machine Learning in NLP

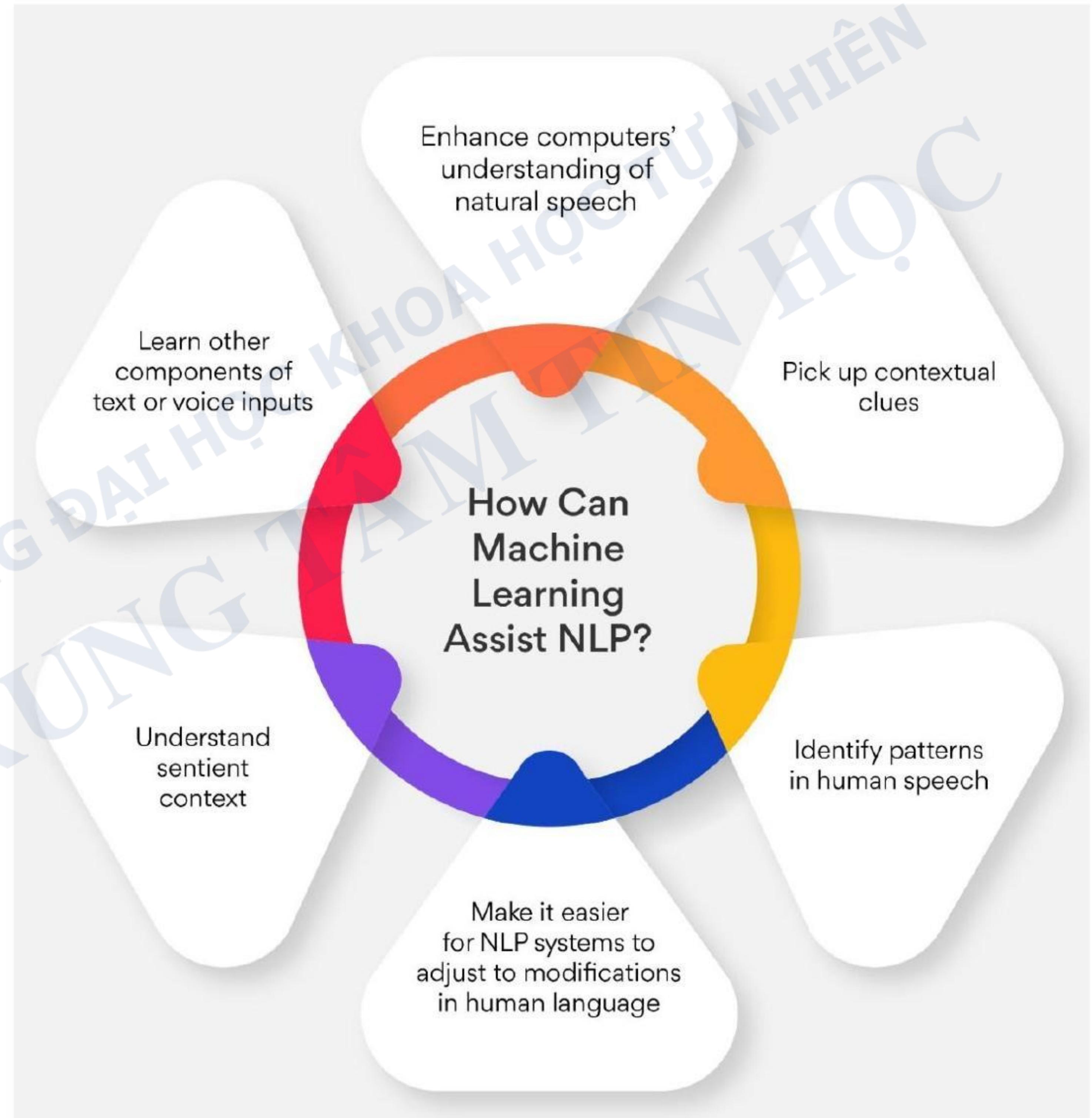
Machine Learning và NLP



Machine Learning và NLP



Các ứng dụng của
Machine Learning
trong Xử lý ngôn ngữ
tự nhiên.





I. Tổng quan Natural Language Processing

II. Chức năng của NLP

III. Natural Language Processing Levels

IV. Cấu trúc NLP

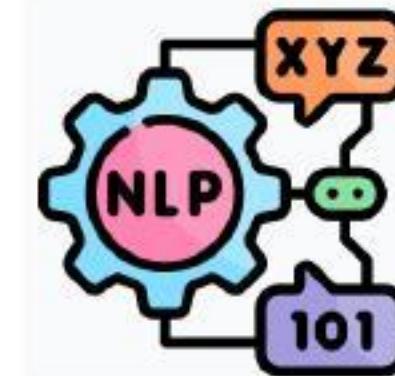


Natural Language Processing Levels

NLP có 7 level cơ bản để hiểu và rút ra ý nghĩa từ một văn bản.



Về âm vị: Phát hiện các từ thông qua âm thanh và lời nói.



Natural Language Processing Levels

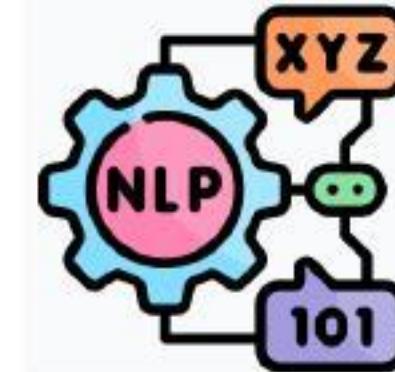
NLP có 7 level cơ bản để hiểu và rút ra ý nghĩa từ một văn bản.



Về hình thái từ: nghiên cứu các từ được xây dựng từ những ý nghĩa nhỏ hơn.

Ví dụ từ **rabbit** có hình thái là 1 trong khi từ **rabbits** có 2 (thêm chữ 's' biểu thị khái niệm số ít và số nhiều).

Xử lý những từ nhỏ nhất có truyền đạt ý nghĩa, các hậu tố và tiền tố.



Natural Language Processing Levels

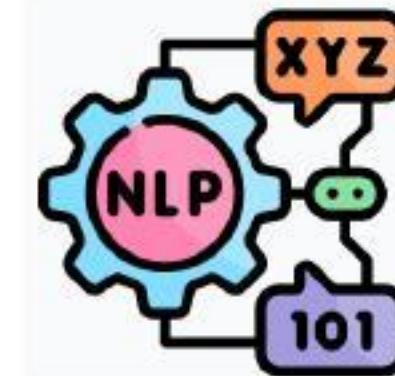
NLP có 7 level cơ bản để hiểu và rút ra ý nghĩa từ một văn bản.



Về từ vựng: nghiên cứu các từ theo ý nghĩa từ vựng và part of speech (POS) của từ đó.

Sử dụng từ vựng là tập hợp các từ vị.

Từ vị là đơn vị cơ bản của từ vựng, có ý nghĩa từ vựng.



Natural Language Processing Levels

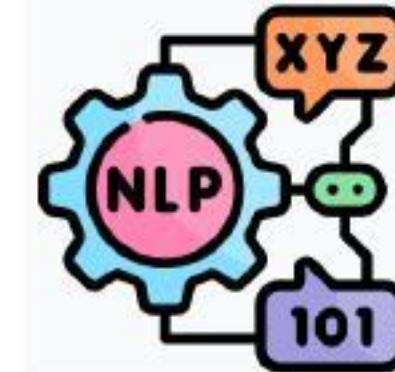
NLP có 7 level cơ bản để hiểu và rút ra ý nghĩa từ một văn bản.



Về cú pháp: ngữ pháp và cấu trúc câu.

Nghiên cứu các mối quan hệ giữa các từ.

POS tagging output của lexical level có thể được sử dụng ở syntactic level để nhóm các từ trong các cụm từ và mệnh đề.



Natural Language Processing Levels

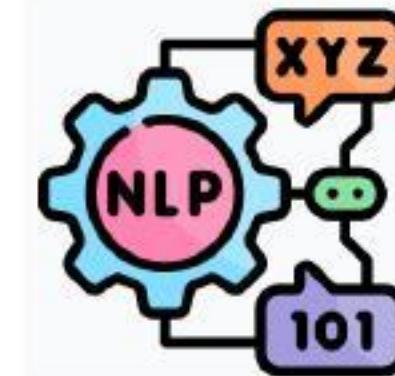
NLP có 7 level cơ bản để hiểu và rút ra ý nghĩa từ một văn bản.



Về ngữ nghĩa: nghiên cứu ý nghĩa của từ và câu có liên quan đến cấu trúc ngữ pháp.

Có hai cách tiếp cận khác nhau:

- 1) Syntax driven semantic analysis
- 2) Semantic grammar

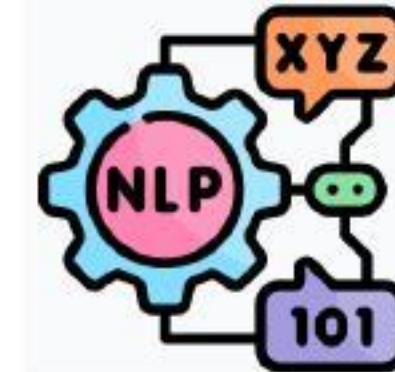


Natural Language Processing Levels

NLP có 7 level cơ bản để hiểu và rút ra ý nghĩa từ một văn bản.



Về diễn ngôn: xác định cấu trúc của các loại văn bản khác nhau.
Có 2 loại diễn ngôn:
1) Anaphora resolution
2) Discourse/ text structure recognition



Natural Language Processing Levels

NLP có 7 level cơ bản để hiểu và rút ra ý nghĩa từ một văn bản.



Về ngữ dụng: nghiên cứu mức độ ảnh hưởng của việc sử dụng kiến thức và hiểu biết thực tế đến ý nghĩa nội dung đang được truyền đạt.

Ngữ dụng xác định ý nghĩa của các từ và cụm từ dựa trên cách ngôn ngữ được sử dụng để giao tiếp.

GIỚI THIỆU NLP



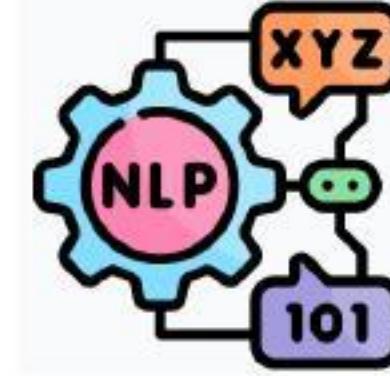
I. Tổng quan Natural Language Processing

II. Chức năng của NLP

III. Natural Language Processing Levels

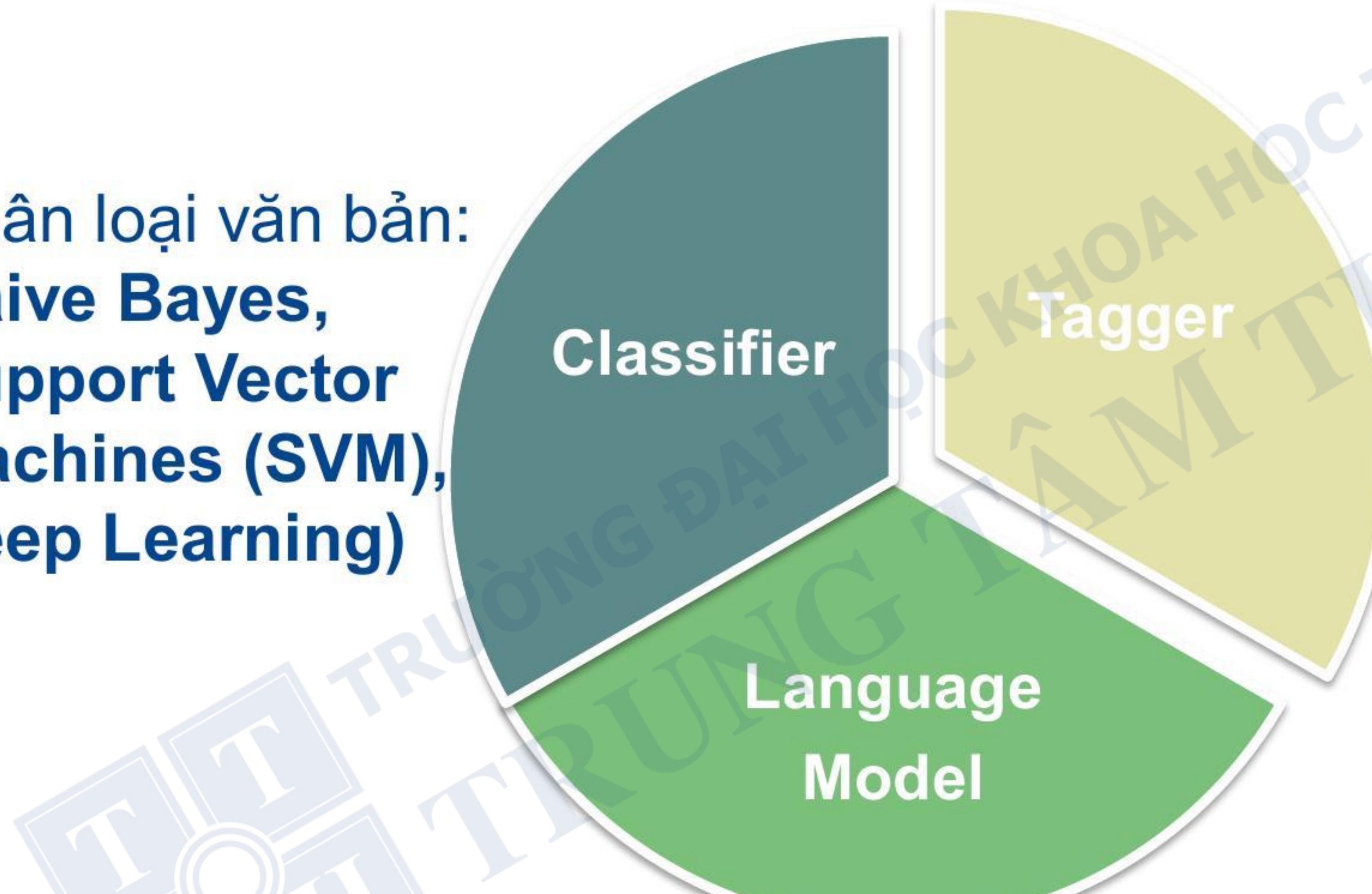
IV. Cấu trúc NLP

Cấu trúc Natural Language Processing



Một hệ thống NLP có các thành phần chính sau:

Phân loại văn bản:
Naive Bayes,
**Support Vector
Machines (SVM),**
Deep Learning)



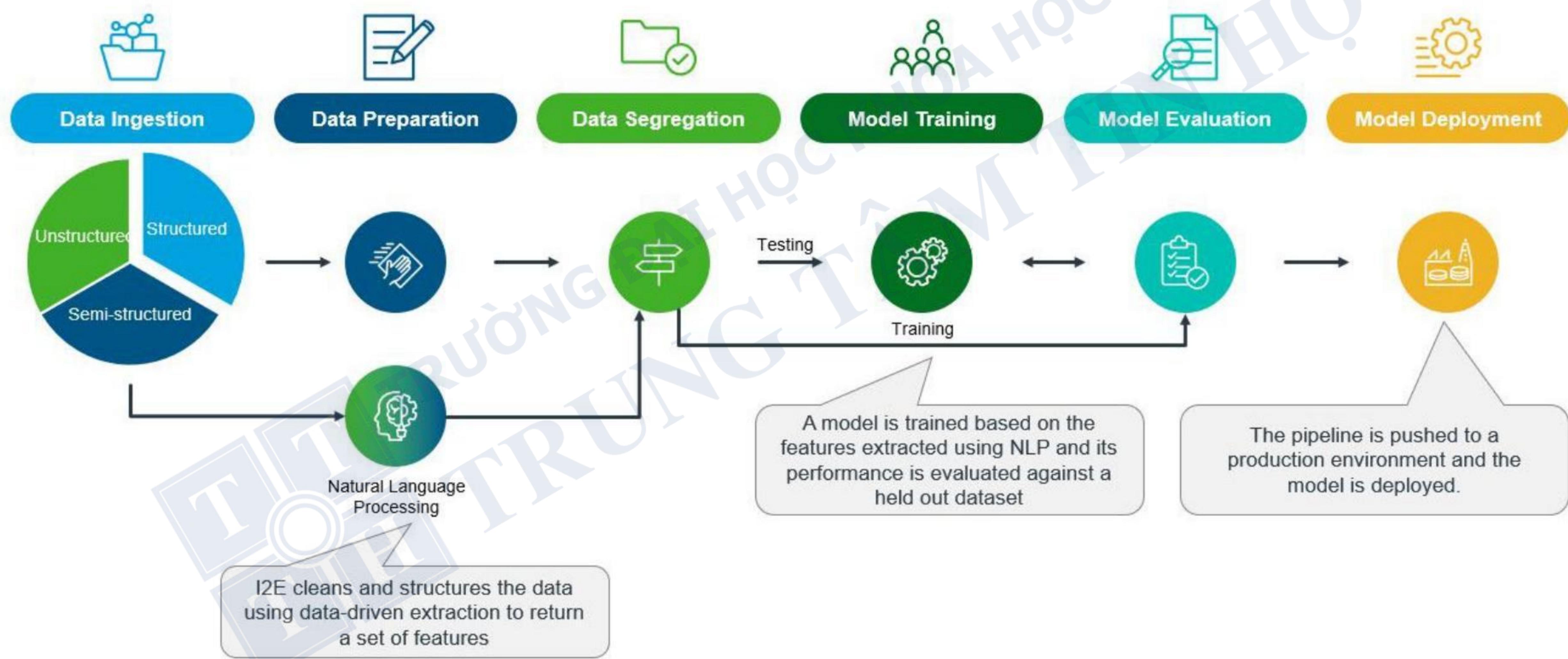
Tagging (POS tags,
named entities, ...):
**Hidden Markov
Models ,
Conditional
Random Fields**

Mô hình ngôn ngữ: **N-gram,**
Recurrent Neural Network,
Transformer Model.

Cấu trúc Natural Language Processing



Quy trình hệ thống NLP



Code Demo



→ Truy cập vào:

[LINK](#)



Q&A

