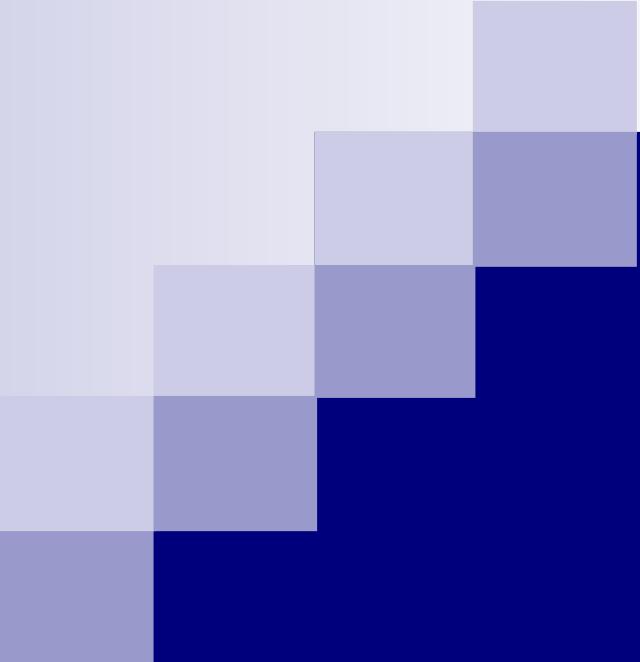


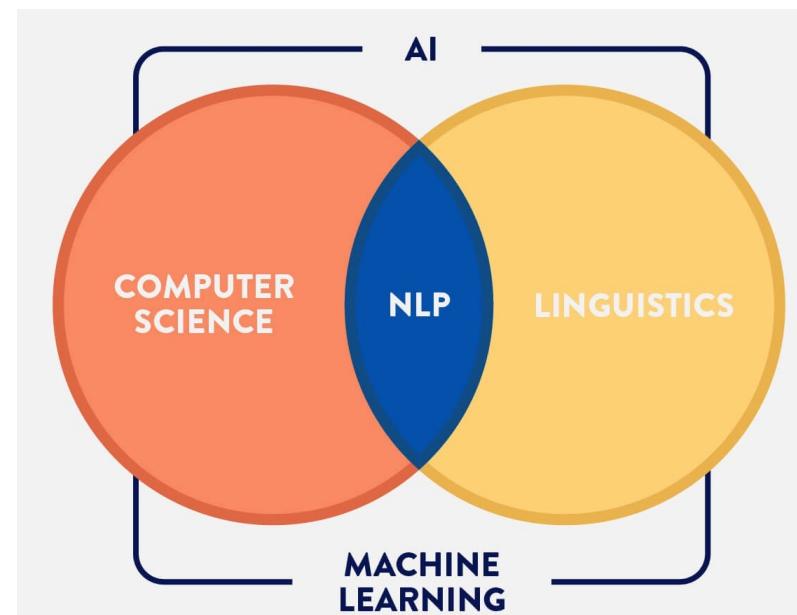
COMP5423 Natural Language Processing



Introduction to Natural Language Process (Part 2)

Introduction to NLP

■ Technologies for NLP



NLP is also known as computational linguistics (CL).

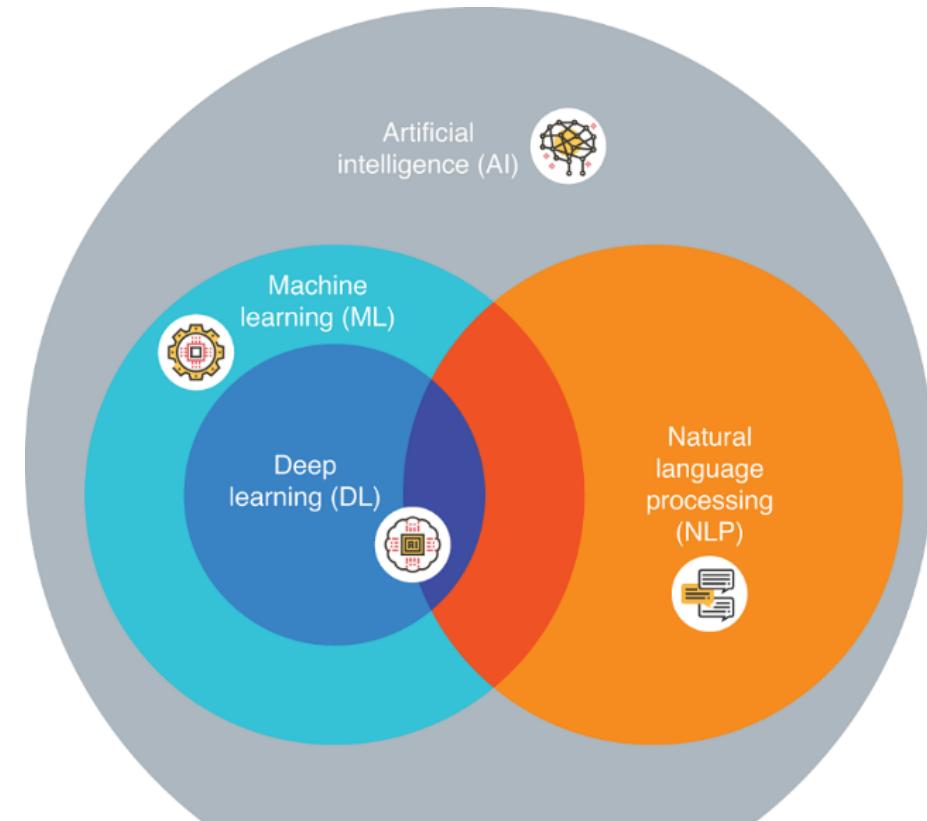
Natural Language Processing (NLP) is a interdisciplinary field of **linguistics**, **computer science** and **artificial intelligence**, concerned with the interactions between computers and human language. [Wikipedia]

Introduction to NLP

■ Technologies for NLP

Artificial intelligence (AI) refers to the simulation of human intelligence in machines that are programmed to think and act like humans.

Natural Language Processing (NLP) is a subfield of AI which has close ties with Machine Learning.



Cognitive scientists have long debated whether language and cognition are separate mental faculties, or whether language emerges from general cognitive abilities.

Introduction to NLP

■ Technologies for NLP

□ Text Classification

- News Categorization
- Entity Relation Classification
- Sentiment, Emotion Classification
- Email Spam, Fake News, Rumor Detection

□ Text Ranking

- Information Retrieval
- Extractive Summarization
- Retrieval-based Response Generation

Introduction to NLP

■ Technologies for NLP

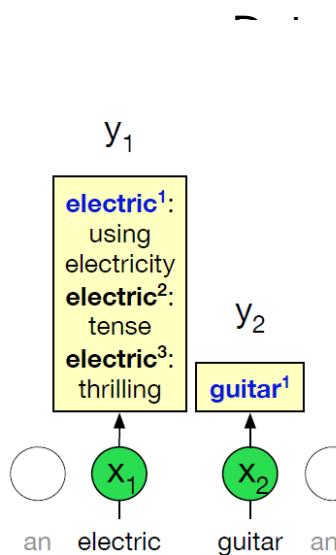
- Text Sequence Labeling
 - Part-of-Speech Tagging
 - Semantic Role Labelling
 - Named Entity Extraction
- Sequence-to-Sequence Text Generation
 - Machine Translation
 - Response Generation
 - Abstractive Summarization
- Syntactic, Semantic, Discourse Parsing

Introduction to NLP

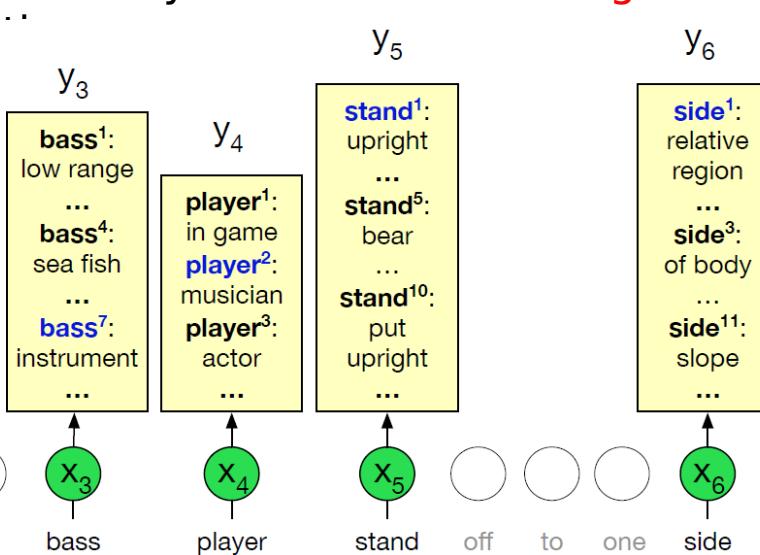
Text Classification

Fundamental NLP Tasks

Sentence Boundary



Sense Disambiguation



Sentiment/ Emotion Classification

Entity Type Classification

The United States President Trump was born in the borough of Queens in New York City.

NLP Application Tasks

Email Spam Classification

Fake News Classification

News Categorizati on

Relation Extraction

Features: (Context) Word, POS, Word Shapes (e.g., Xx, x), Phrase (NE) Types, Syntactic Path, Lexicon

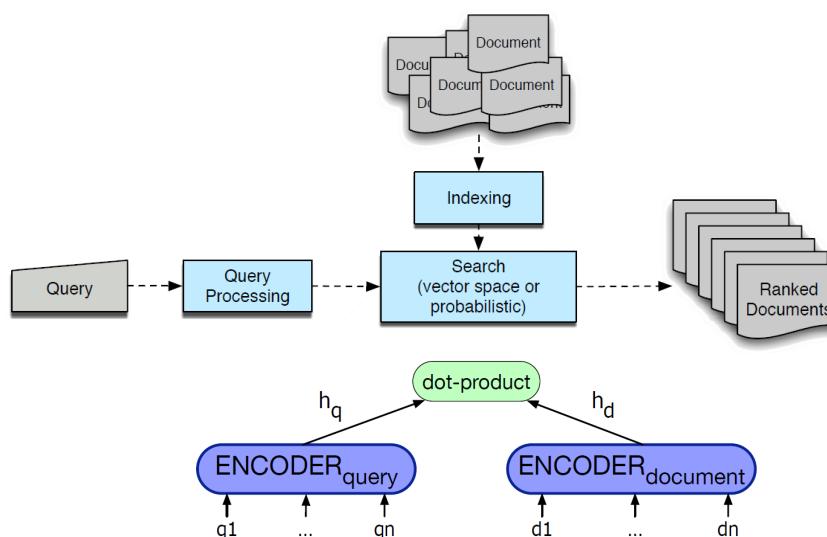
Classifiers: Naïve Bayes, Decision Tree, Random Forests, SVM, Logistic Regression, RNN, CNN, BERT,

Introduction to NLP

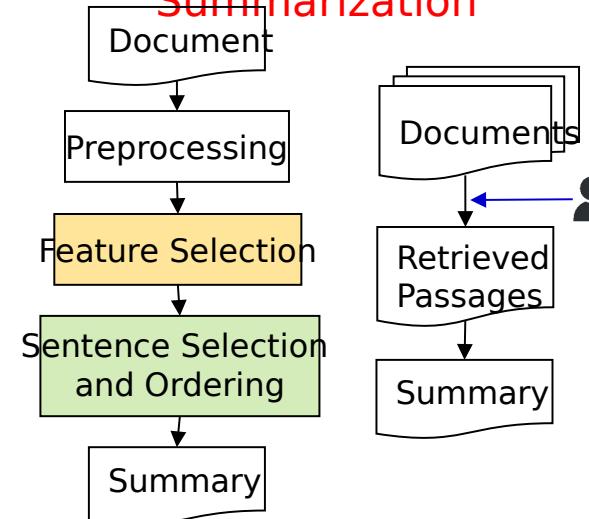
Text Ranking

NLP Application Tasks

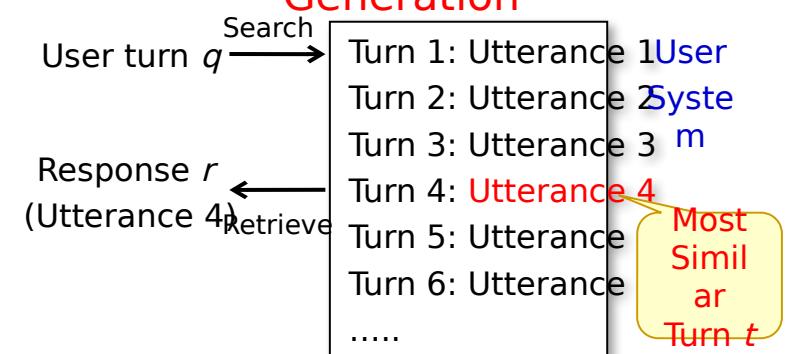
Information Retrieval



Extractive Summarization



Retrieval-based Response Generation



Features: Word, Position,....

Retrieval Models: Vector Space Model, Probabilistic Retrieval Model, Language Model, Dense Retrieval Model

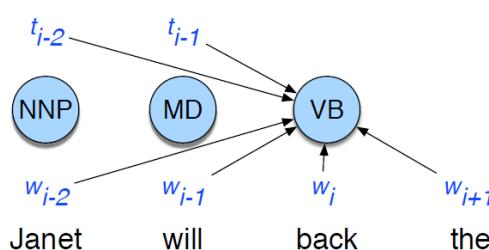
Introduction to NLP

Sequence Labeling

Fundamental NLP Tasks

NLP Application Tasks

Part-of-Speech Tagging



Named Entity Tagging (BIO Annotation Scheme)

Donal	d	Trum	p	wa	s	a	preside	nt	of	USA
	B-PER	I-PER	O	O	O	O	O	O	B-LOC	
Donal	d	Trum	p	was	a	preside	nt	of	USA	
Noun	Noun	Ver	b	Articl	e	Noun	PREPO	S	Nou	

Answer Extraction

In the box, you get the camera itself inside a mounting case, a warranty card, user manual, and many accessories. It does require a microSD card, as there is no internal storage. Upon first use, I found the camera to fell pretty nice...

Question: Do I need SD card?
Answer: require a microSD card

Sequential Classifiers: HMM, Conditional Random Fields (CRF), RNN, BERT, ...

Introduction to NLP

Sequence-to-Sequence Generation

NLP Application Tasks

Machine Translation

Abstractive Summarization

Response Generation

Answer Generation

Where was Alan Turing born?

Alan Turing was a British computer scientist. Born in Maida Vale, London...

Generative seq2seq model

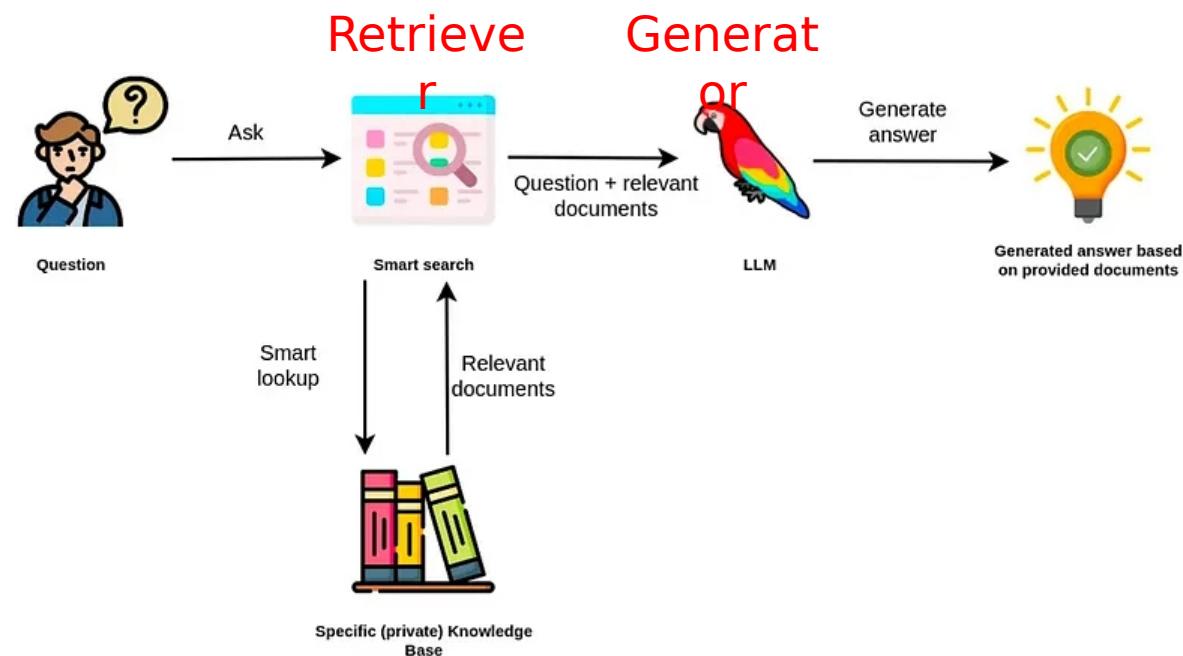
Maida Vale, London

Retrieval-Augmented Generation

Sequence-to-Sequence Models: BART, T5, ...

Introduction to NLP

Sequence-to-Sequence Generation

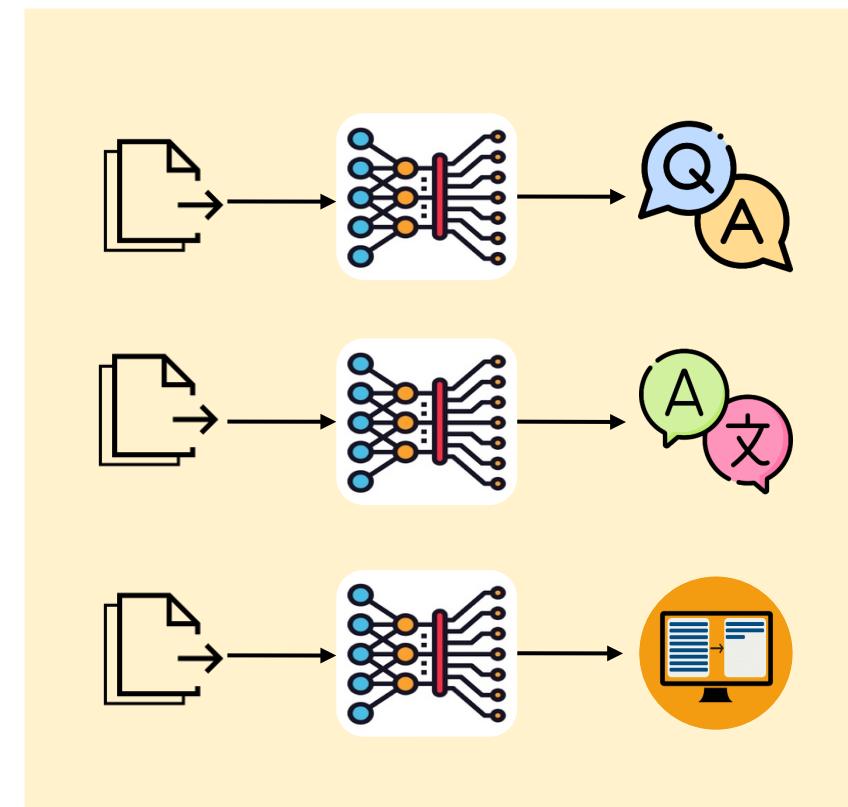
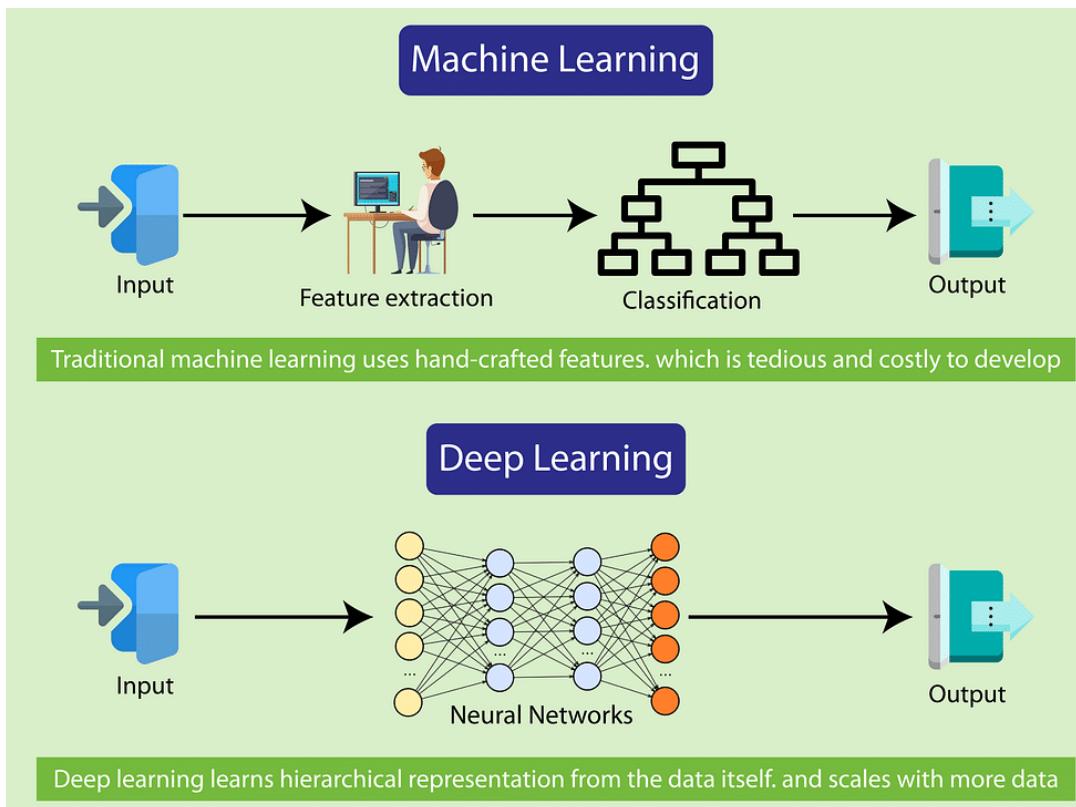


Retrieval-Augmented Generation
(RAG)

Introduction to NLP

■ Technologies for NLP

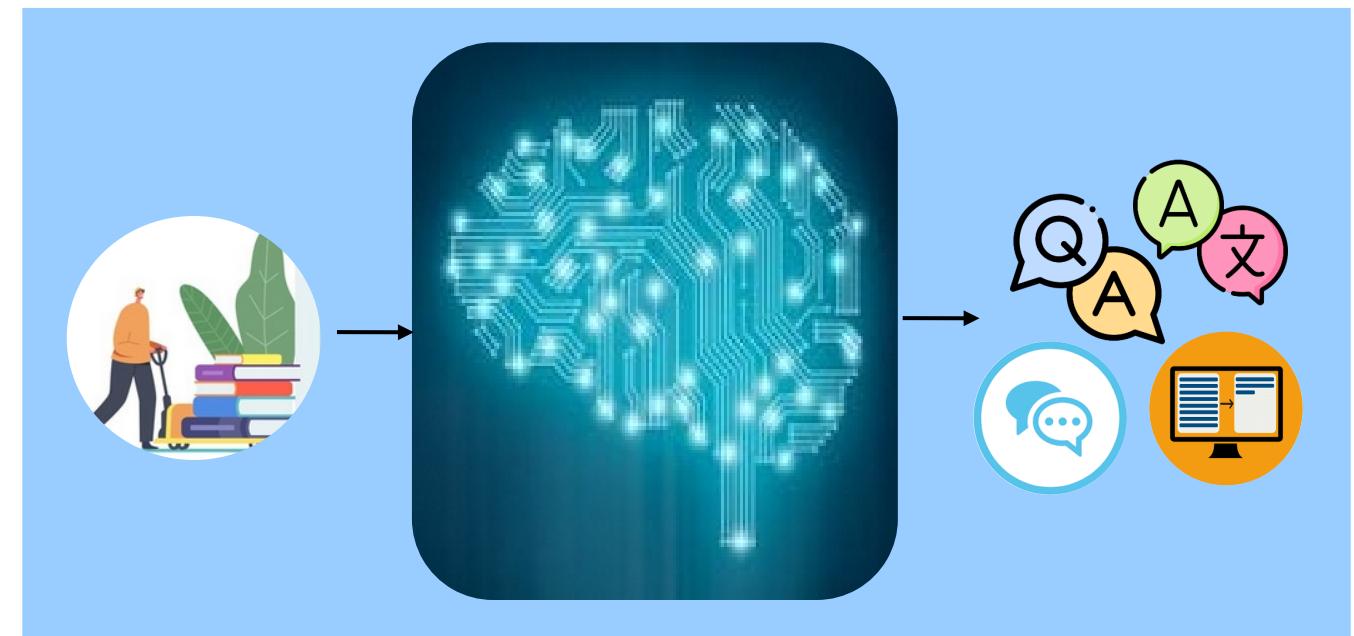
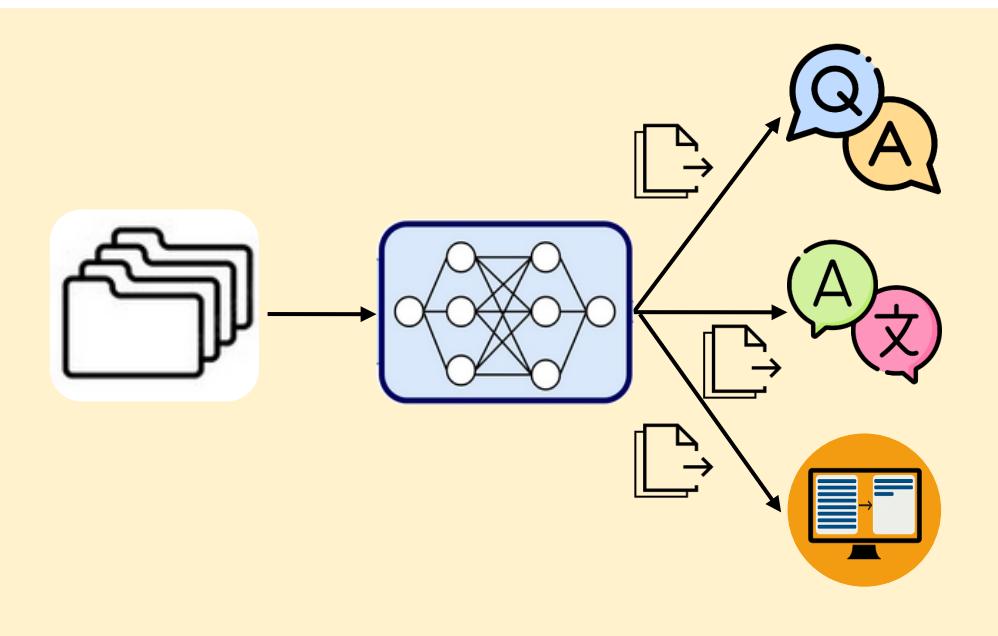
□ From Machine Learning to Deep Learning



Introduction to NLP

■ Technologies for NLP

- From Pre-Trained Language Models to Large Language Models



Introduction to NLP

■ History of NLP

- Phase 1: Symbolic NLP (1950s – 1990s)
 - Models of Grammars (Syntactics)
 - Logical Form (Semantics and Reasoning)
 - (Linguistic) Rule-based Approaches
- Phase 2: Statistical NLP (1990s – 2010s)
 - Machine Learning based Approaches (Supervised)
 - Feature Engineering

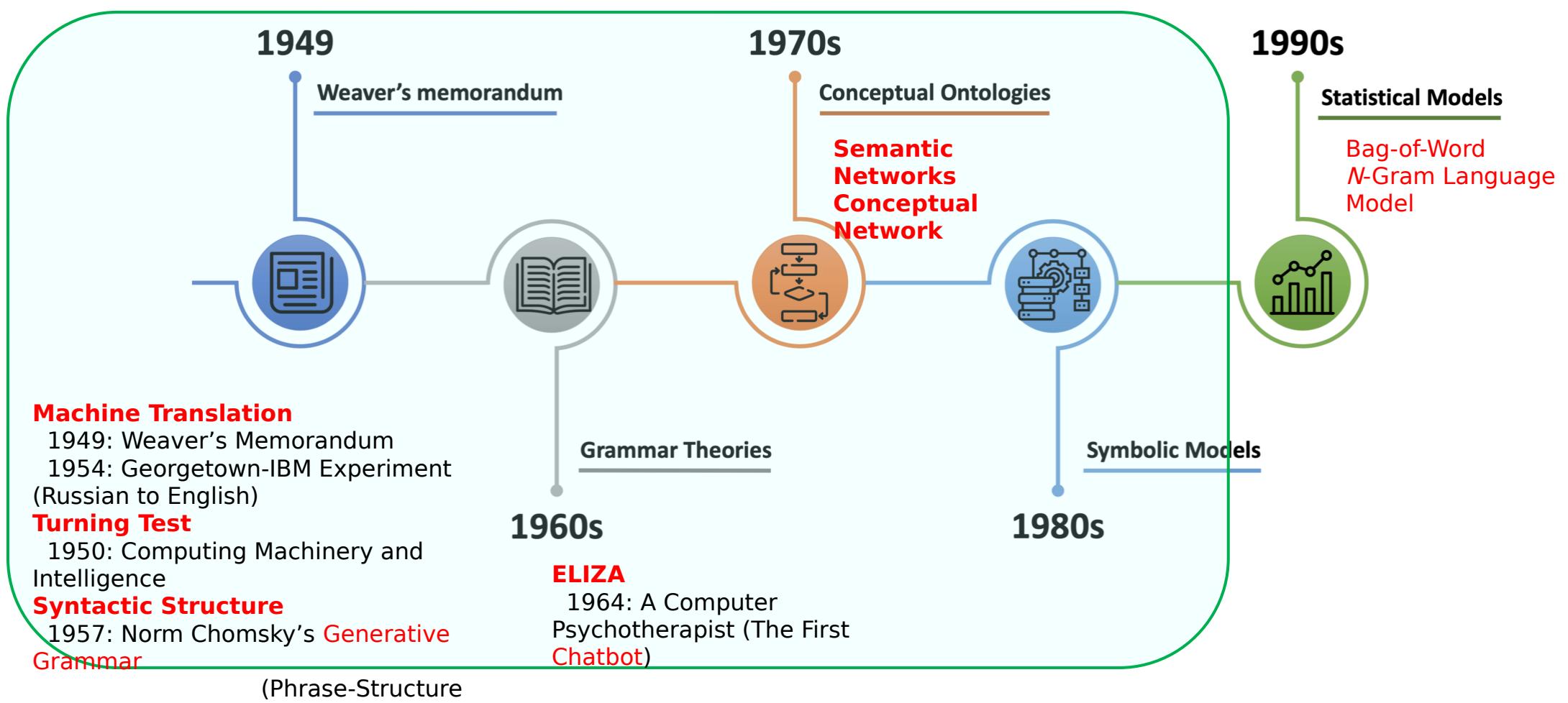
Introduction to NLP

■ History of NLP

- Phase 3: Neural NLP (2010s – Present)
 - Neural Network Models
 - Deep Learning based Approaches
 - Representation Learning (Self-Supervised)

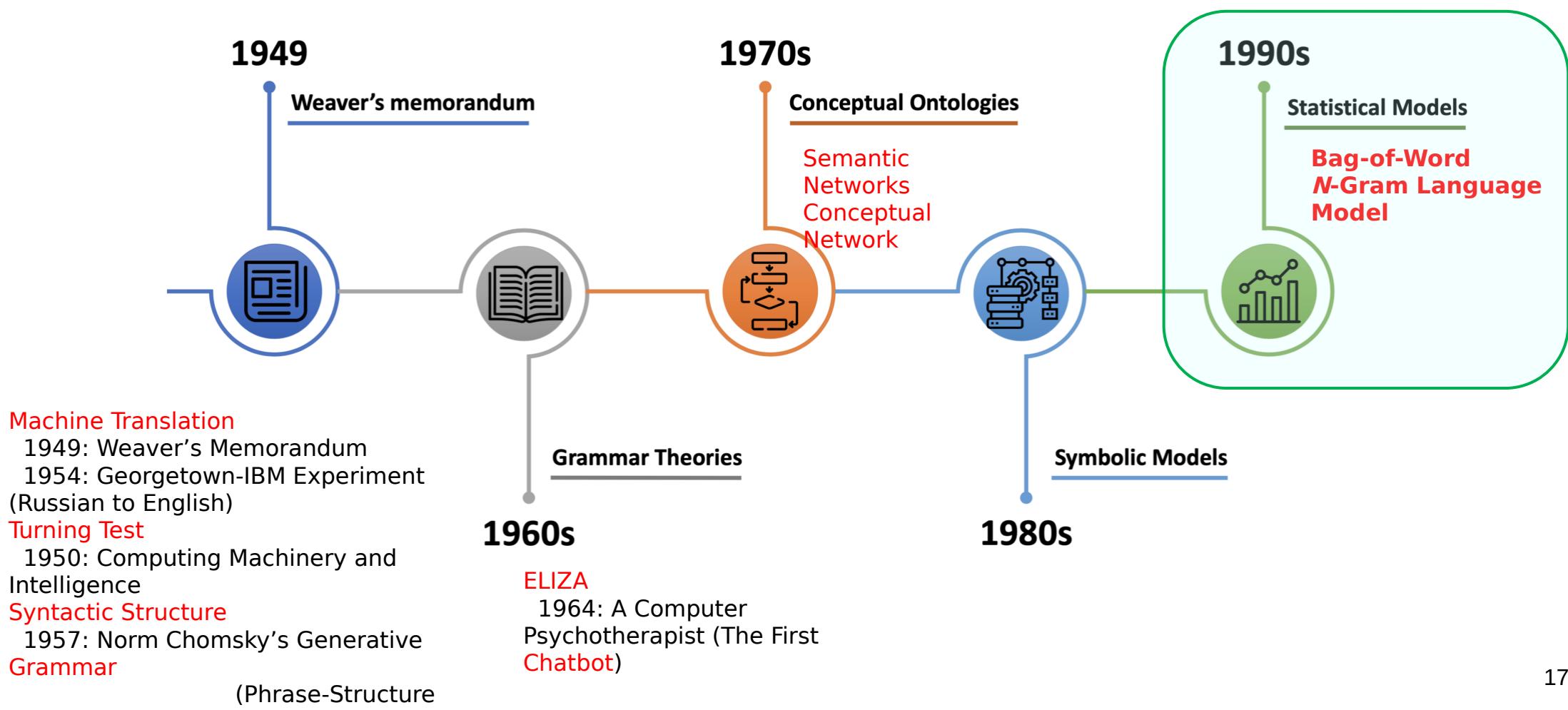
Introduction to NLP

■ History of NLP



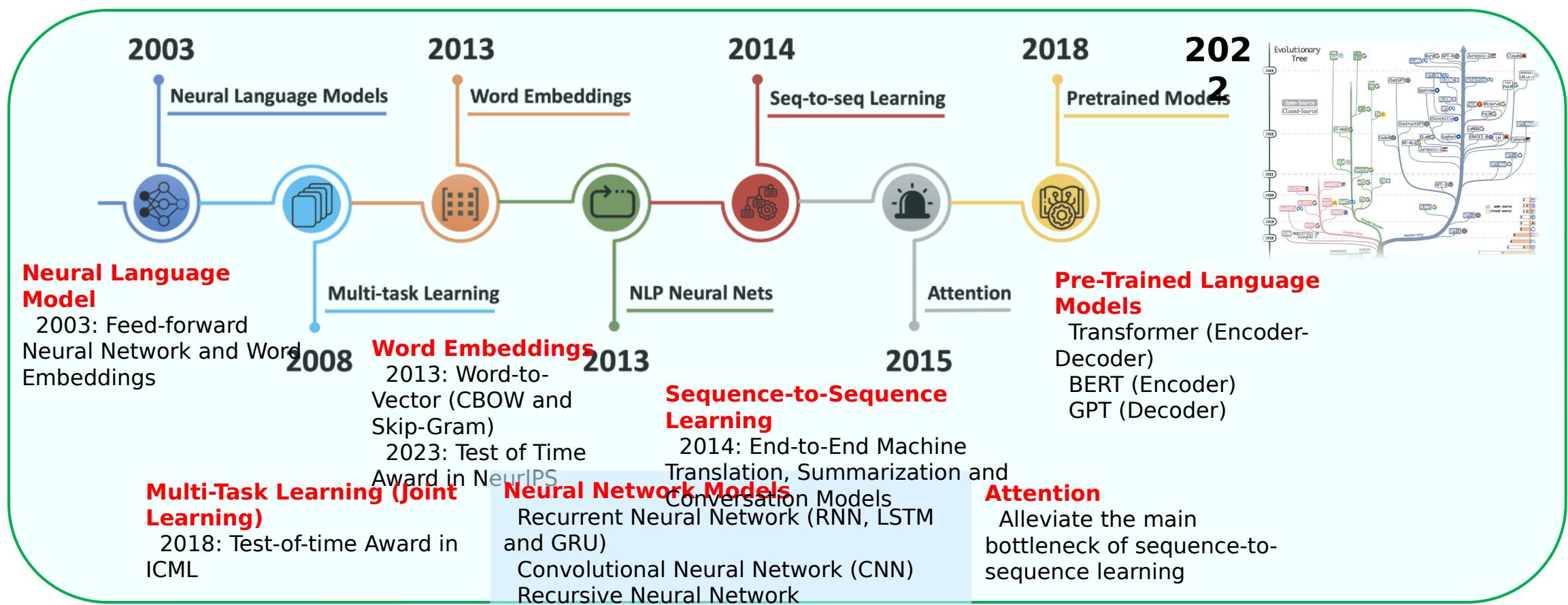
Introduction to NLP

■ History of NLP



Introduction to NLP

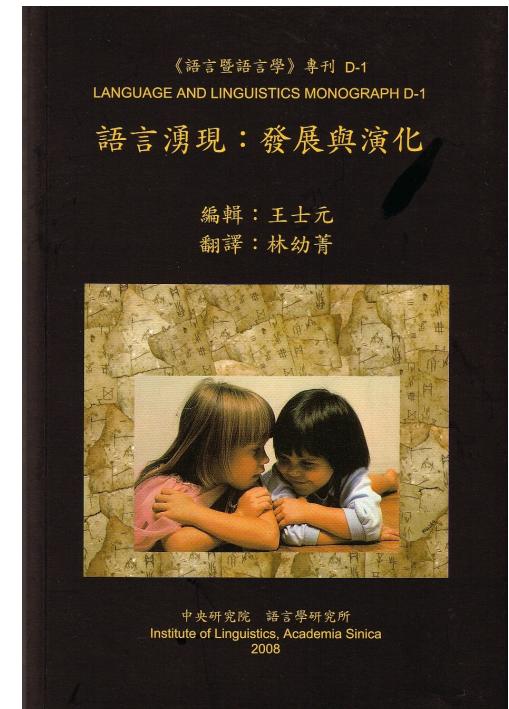
■ History of NLP



Introduction to NLP

■ Challenges in NLP

- Language is Flexible: Open-ended Vocabulary,
Flexible Combination, No Fixed Grammar
- Language is Ambiguity (at Different Levels)



Introduction to NLP

■ Challenges in NLP

Word-Level Grammatical Ambiguity

A word could be used as a verb, noun, or adjective.
earnings growth took a back/JJ seat
a small building in the back/NN
a clear majority of senators back/VBP
the bill

Dave began to back/VB toward the door
enable the country to buy back/RP
about debt
I was twenty-one back/RB then

Word-Level Semantic Ambiguity

An word may have multiple meanings (aka. senses)



.... a quiet animal like a **mouse**.



.... a **mouse** controlling a computer system in 1968.

Introduction to NLP

■ Challenges in NLP

Word Segmentation Ambiguity

There are more than one possible segmentation
for an input character string.

日文章魚怎么說



日文章魚怎麼說

How to say **octopus** in
Japanese

日文章魚怎麼說

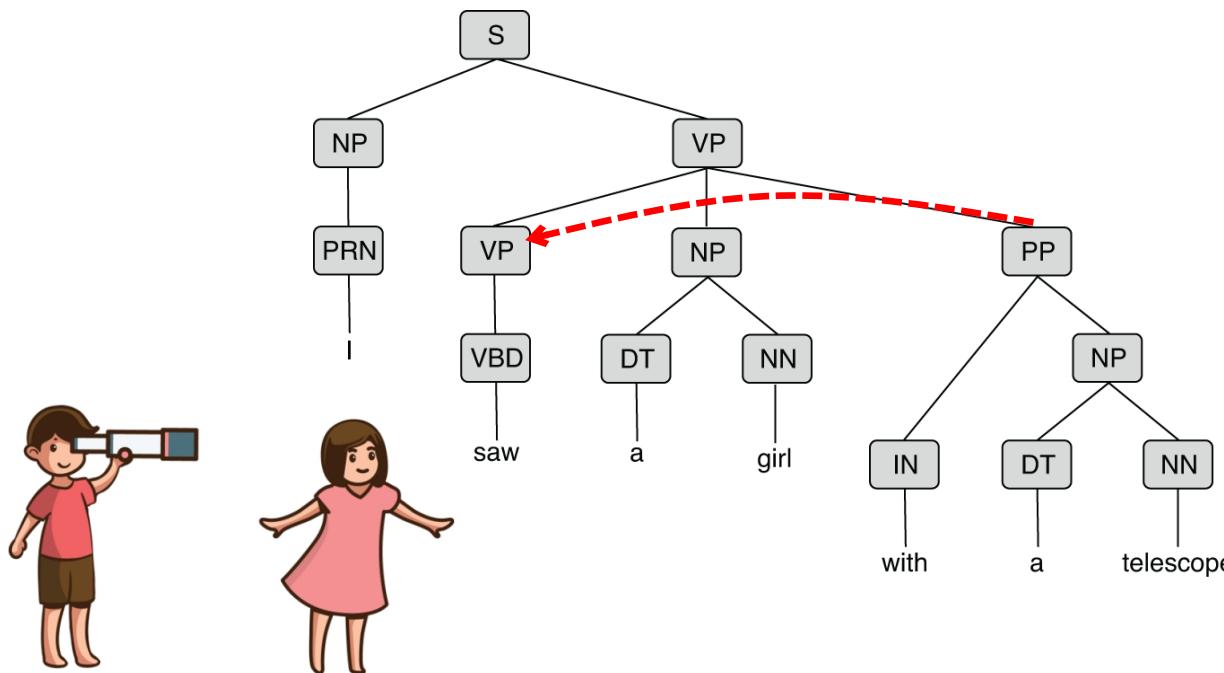
How fish says Japanese
article

Introduction to NLP

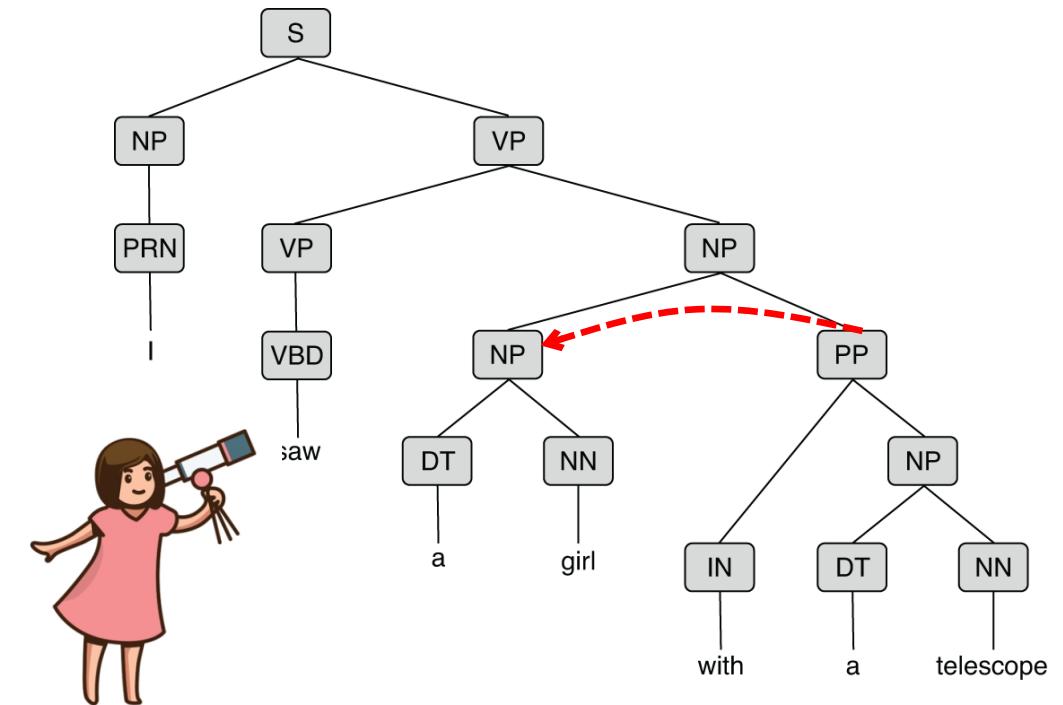
■ Challenges in NLP

Sentence-Level Syntactic Ambiguity

An preposition phrase (PP) can attached to a verb phrase (VP) or a noun phrase (NP).



I saw a girl with a telescope.



I saw a girl with a telescope.

Introduction to NLP

■ Challenges in NLP

Coreference Ambiguity

A pronoun may refer to more than one preceding entity

The firemen arrived after the police officer because they were coming from so far away.

This diagram illustrates coreference ambiguity. The word 'they' is highlighted in yellow and points to two different entities: 'firemen' and 'police officer'. A bracket above the sentence spans both entities, indicating that the pronoun could refer to either or both of them.

The firemen arrived after the police because they were coming from so far away.

This diagram illustrates coreference ambiguity. The word 'they' is highlighted in yellow and points to two different entities: 'firemen' and 'police'. A question mark above the sentence indicates that it is ambiguous which entity the pronoun refers to.

Introduction to NLP

■ Challenges in Generative AI

□ Hallucinations refers to the generation of outputs that may sound plausible but are either factually incorrect or unrelated to the given context.



When did Leonardo da Vinci paint the Mona Lisa?



Leonardo da Vinci painted the Mona Lisa in 1815.

(Incorrect: The Mona Lisa was painted between 1503 and 1506, or perhaps continuing until 1517.)

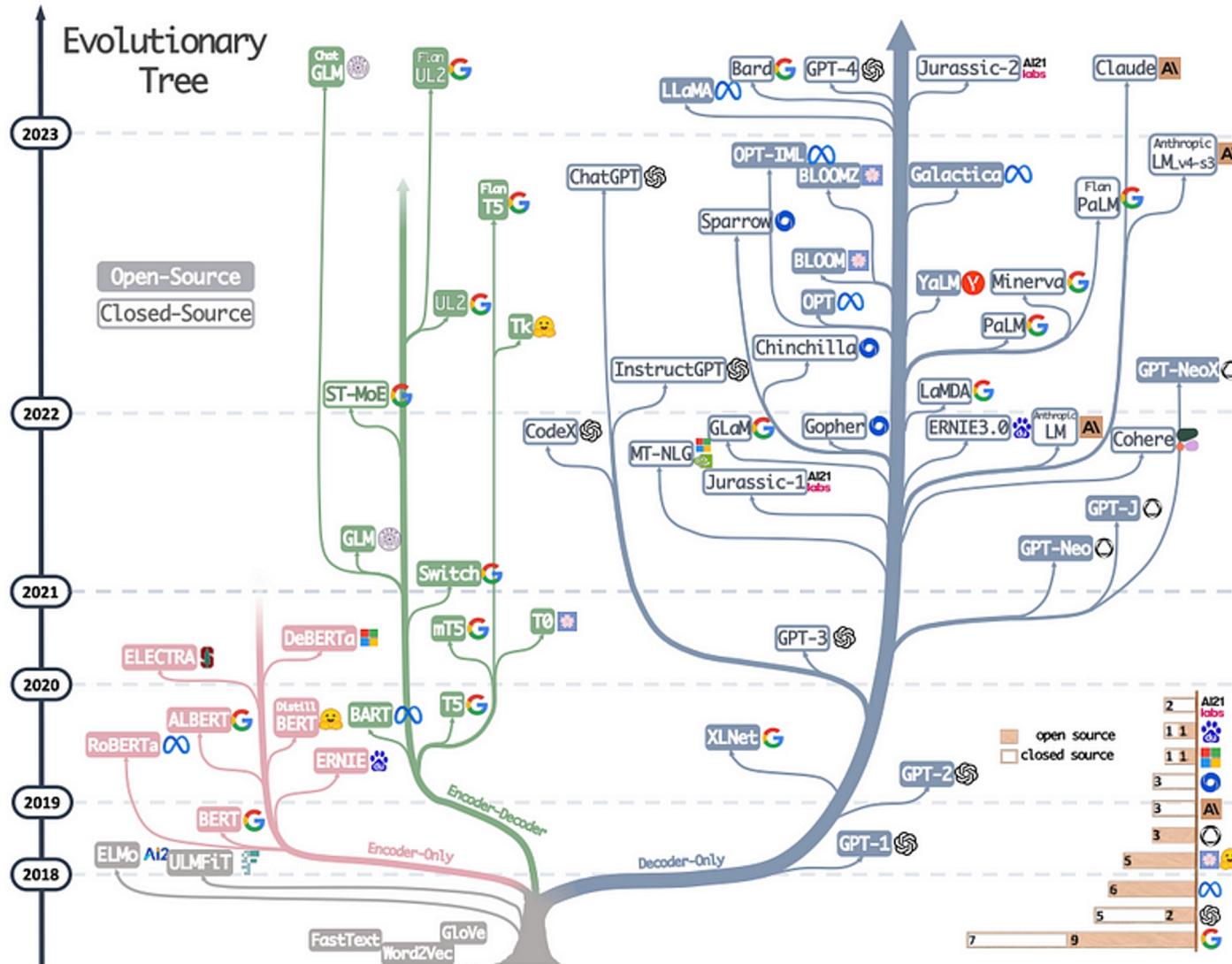


2025
HAPPY NEW YEAR
YEAR OF THE SNAKE



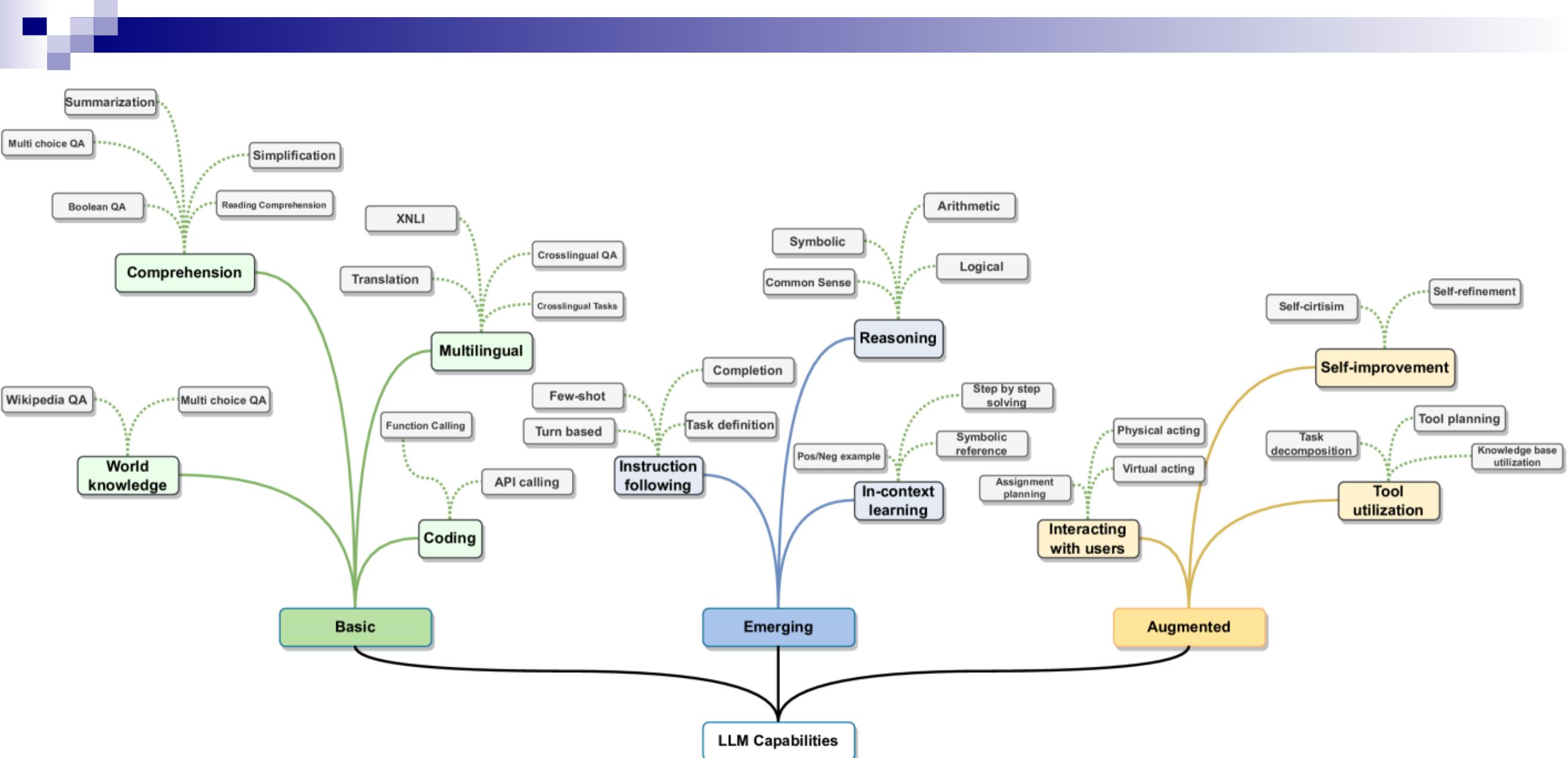


Thank you

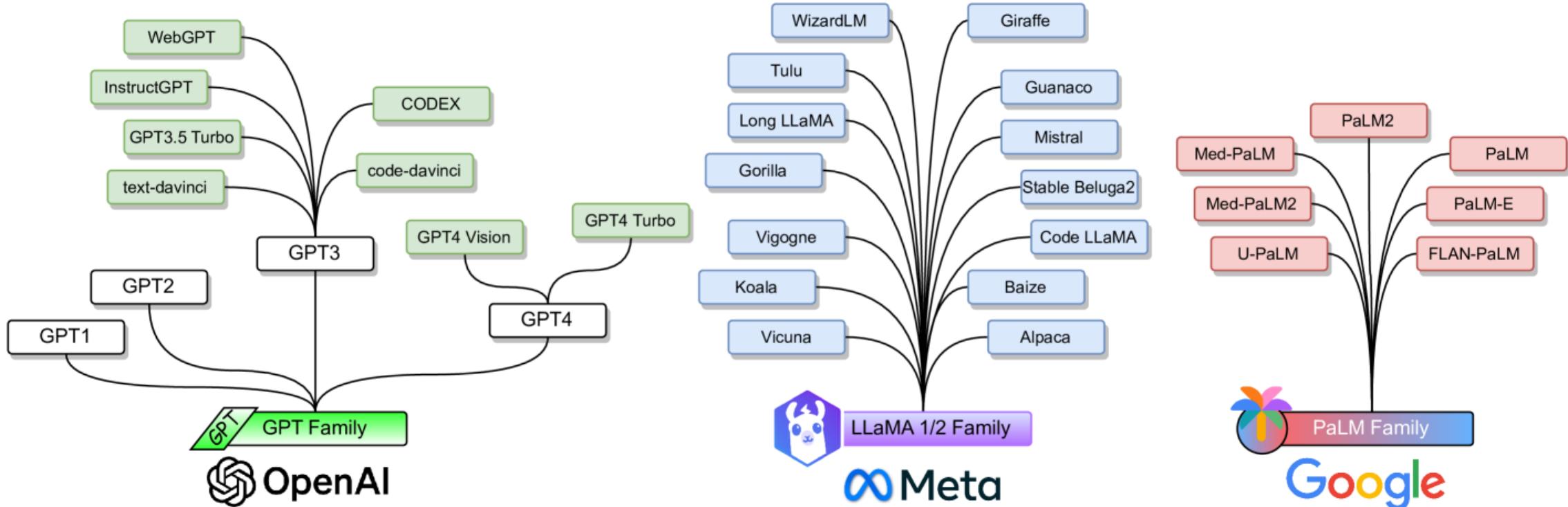


Large Language Model (LLM) Evolutionary Tree
(Source:

Harnessing the Power of LLMs in Practice: A Survey on ChatGPT and



Large Language Model (LLM) Capabilities (Source:
[Large Language Models: A Survey](#))



Large Language Model (LLM) Capabilities (Source:
[Large Language Models: A Survey](#))