

Advanced Natural Language Processing

정상근

Sangkeun Jung

Introduction

NATURAL LANGUAGE PROCESSING

What is

Natural Language?

What about?

Artificial Language

Machine Language

Formal Language

Programming Language

Time



스 · 야 · 하 · 시 · 셔 · 이 · 런 · 수 · 달 · 나
를 · 어 · 니 · 러 · 도 · 니 · 전 · 못 · 아 · 랫
들 · 엇 · 라 · 퍼 · 무 · 르 · 초 · 디 · 문 · 말
쥬 · 비 · 내 · 디 · 춤 · 그 · 로 · 아 · 쥬 · 쑥
쥬 · 너 · 이 · 몬 · 내 · 저 · 어 · 니 · 와 · 미
를 · 겨 · 를 · 흥 · 제 · 흥 · 린 · 홀 · 르 · 등
밍 · 새 · 윙 · 노 · 브 · 배 · 린 · 쑥 · 서 · 귀
그 · 로 · 후 · 미 · 들 · 이 · 생 · 이 · 르 · 에

Aspects of NL

Place



Technology



Technology



Gandalf the Grey checked into Rivendell with 6 other people.

December 25, 3018 via Great Eagle - - Like - Comment



Gandalf the Grey So it's settled then. I'll talk to the Eagles and we will fly the ring into Mt. Doom

December 25, 3018 · Like



Boromir Whoa whoa whoa, one does not simply parachute into Mordor

December 25, 3018 · Like



Samwise Gamgee Why not?

December 25, 3018 · Like



Gimli Yeah, it sounds pretty easy

December 25, 3018 · Like



Legolas I can't think of any other way to do it

December 25, 3018 · Like



Frodo Baggins How else are we going to get there...walk?

December 25, 3018 · Like



Samwise Gamgee LMFAO!!!

December 25, 3018 · Like

For example:



Blablabla500 reviewed



"Rubbish for a supposedly prest..."



Dorothy M reviewed



"Never again"



Linda B reviewed



"Empty venue, still poor service!!"

Format

[Poem]

새해라, 기쁜 새해라
기쁜 일 자주 생겨 기뻐하였으면,
풍년 들고 과실이 잘 익어서
태평 시절에 한가한 사람 되었으면.

新年喜新年 喜事喜頻頻
(신년희신년 회사희빈빈)
田穀豐登園果好 太平時節作閑人
(전곡풍등원과호 태평시절작한인)

새해라, 좋은 새해라
좋은 일 자랑할 만하면 좋겠네.
앞마을 노랫가락 뒷마을에 이어지고
앞집 늙은이 뒷집 늙은이와 부유함을 다투기를.

新年好新年 好事好堪誇
(신년호신년 호사호감과)
南里人歌廣北里 東家翁富闊西家
(남리인가광북리 동가옹부투서가)

새해라, 즐거운 새해라
즐거운 일 더욱 넉넉하여 즐겁기를.
집집마다 찢은 곡식 창고에 넘쳐나고
사람마다 지은 옷 께짝에 가득했으면.

新年樂新年 樂事樂更怏
(신년낙신년 낙사낙경협)
家家春粟溢困倉 人人製衣盈箱篋
(가가용속일균창 인인제의영상협)
-박세당, '춘첩(春帖)'

Aspects of NL

Format

[Letter]



Aspects of NL

Format

[Book]



Aspects of NL

Age



Language?



- ① Who is the sender of NL?
- ② Who is the receiver of NL?
- ③ Does the sender know the receiver?
- ④ What sender send =? What receiver receive?
- ⑤ Does time matter?
- ⑥ ...



Language = (culture, experience, knowledge,)

Extreme Example





아래의 상황에서 B가 말할 만한 내용은?



[상황]

A: 어제 새벽에 내가 4시 정도에 깨서 부엌에 갔어. 목이 말라서 유리컵에 물을 부어서 마시려다가, 그만 컵을 놓쳐버렸지 뭐야

B:



B가 이 상황에서 말할 만한 내용은 다음과 같을 수 있습니다:



B: "와, 정말 깨어서 물 마시러 갔다가 컵을 놓쳤다니! 다행히 큰 문제가 없었나요? 조심해야겠어요."

ChatGPT 3.5 (2023.09.04)

아래의 상황에서 B가 말할 만한 내용은?

[상황]

A : 어제 새벽에 내가 4시 정도에 깨서 부엌에 갔어. 목이
말라서 유리컵에 물을 부어서 마시러 가다가, 그만 컵을
놓쳐버렸지 뭐야

B :



B: "헉, 그 시간에 정신 없었나 보다! 다행히 다치진 않았어? 유리컵 깨지면 위험한
데."

ChatGPT 4o (2024.08.27)



제가 B의 반응을 "헉, 그 시간에 정신 없었나 보다! 다행히 다치지 않았어? 유리컵 깨지면 위험한데."라고 설정한 이유는 다음과 같습니다:

1. **공감과 위로:** A가 새벽에 유리컵을 떨어뜨리는 경험을 했다는 이야기를 했으니, B는 A의 상황에 공감하면서 놀란 반응을 보이는 것이 자연스럽습니다. "헉"이라는 표현은 A가 겪은 갑작스러운 상황에 대한 놀라움을 표현하며, 이는 A의 경험에 대해 이해하고 있음을 나타냅니다.
2. **안전 확인:** 유리컵이 깨졌을 가능성이 있으므로, A가 다치지 않았는지 확인하는 것은 B의 입장에서 자연스러운 반응입니다. 다치지 않았다는 사실을 확인하려는 질문은 B가 A의 안전을 걱정하고 있음을 보여줍니다.
3. **상황에 대한 언급:** "그 시간에 정신 없었나 보다!"라는 말은 A가 새벽이라는 피곤한 시간에 일어난 일이라는 점을 이해하고, 그 상황을 고려한 공감을 나타내기 위해 추가된 부분입니다. 이는 B가 A의 상황을 이해하고 있다는 것을 더 명확히 해줍니다.

이러한 요소들은 B가 친구 A의 상황을 공감하고, 안전을 염려하는 따뜻한 마음을 표현하기에 적절하다고 판단했습니다.



ChatGPT 4o (2024.08.27)

HS

아래의 상황에서 B가 말할 만한 내용은? [상황] A : 어제 새벽에 내가 4시 정도에 깨서 부엌에 갔어. 목이 말라서 유리컵에 물을 부어서 마시러 가다가, 그만 컵을 놓쳐버렸지 뭐야 B :



이런 상황에서 B가 말할 만한 내용은 다음과 같을 수 있습니다:

"아이고, 그랬구나. 컵이 깨져서 청소하느라 고생했겠다. 다행히 다치지는 않았어? 깨진 유리 조각에 베이기라도 했으면 큰일 날 뻔했네. 그래도 밤중에 물 마시러 가다 그런 일이 생기면 누구나 당황스럽지. 다음부터 조심해야겠다. 혹시 도움이 필요하면 언제든지 나한테 연락해."

이 대답은 친구나 가족 간의 대화처럼 공감과 걱정을 담고 있습니다. B는 A의 경험에 대해 이해를 표현하고, A의 안전을 걱정하며, 위로의 말을 건네고 있습니다. 또한 앞으로 비슷한 상황이 생길 경우를 대비해 조언과 도움을 제안하고 있습니다.

Copy Retry

Claude 3.5 Sonnet (2024.08.27)

HS 왜 저렇게 네가 답했는지 설명해줄래?



네, 제가 그렇게 답변한 이유를 설명해드리겠습니다:

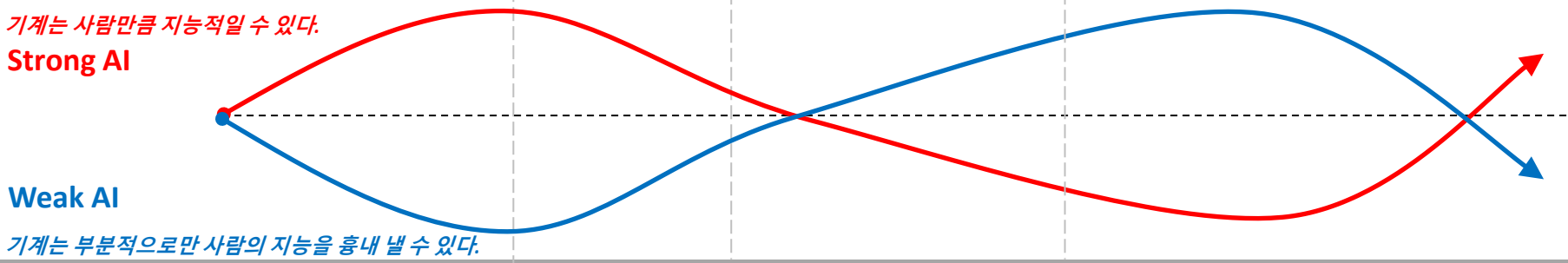
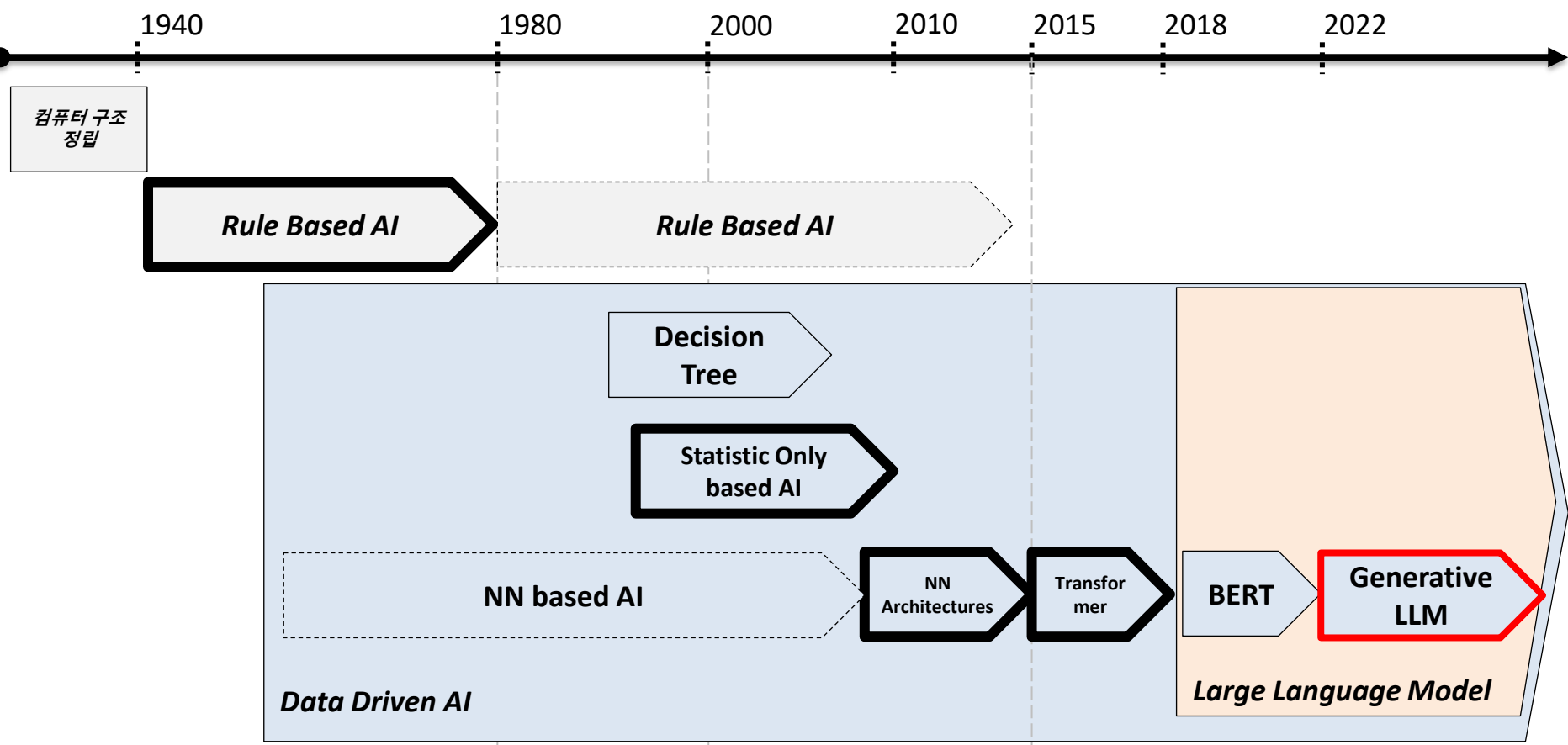
1. 공감과 이해: "아이고, 그랬구나"로 시작하여 **A**의 경험에 대한 이해와 공감을 표현했습니다. 이는 대화 상대방의 감정을 인정하고 공감대를 형성하는 데 중요합니다.
2. 걱정과 안부: "다행히 다치지는 않았어?"라고 물어봄으로써 **A**의 안전에 대한 걱정을 표현했습니다. 이는 친구나 가족 간의 대화에서 흔히 볼 수 있는 반응입니다.
3. 상황에 대한 이해: "그래도 밤중에 물 마시러 가다 그런 일이 생기면 누구나 당황스럽지"라고 말함으로써 **A**의 실수를 정상화하고 위로를 전달했습니다.
4. 조언: "다음부터 조심해야겠다"라고 하여 앞으로의 유사한 상황을 예방하기 위한 간단한 조언을 제시했습니다.
5. 도움 제안: "혹시 도움이 필요하면 언제든지 나한테 연락해"라고 말하며 앞으로도 지지와 도움을 제공할 의사가 있음을 표현했습니다.

이러한 요소들을 포함함으로써, **B**의 대답은 친구나 가족 간의 자연스러운 대화를 반영하고 있습니다. 이는 상대방의 감정을 고려하고, 공감하며, 지지를 보내는 따뜻한 대화 방식입니다.

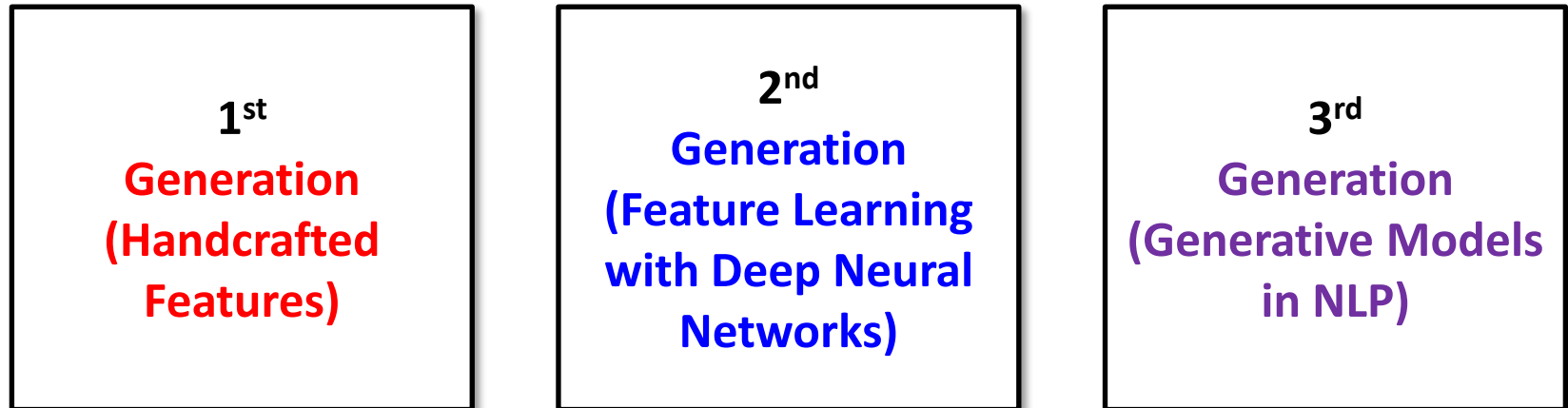
Copy Retry

Claude 3.5 Sonnet (2024.08.27)

Historical View - Artificial Intelligence (NLP)



NLP Generations

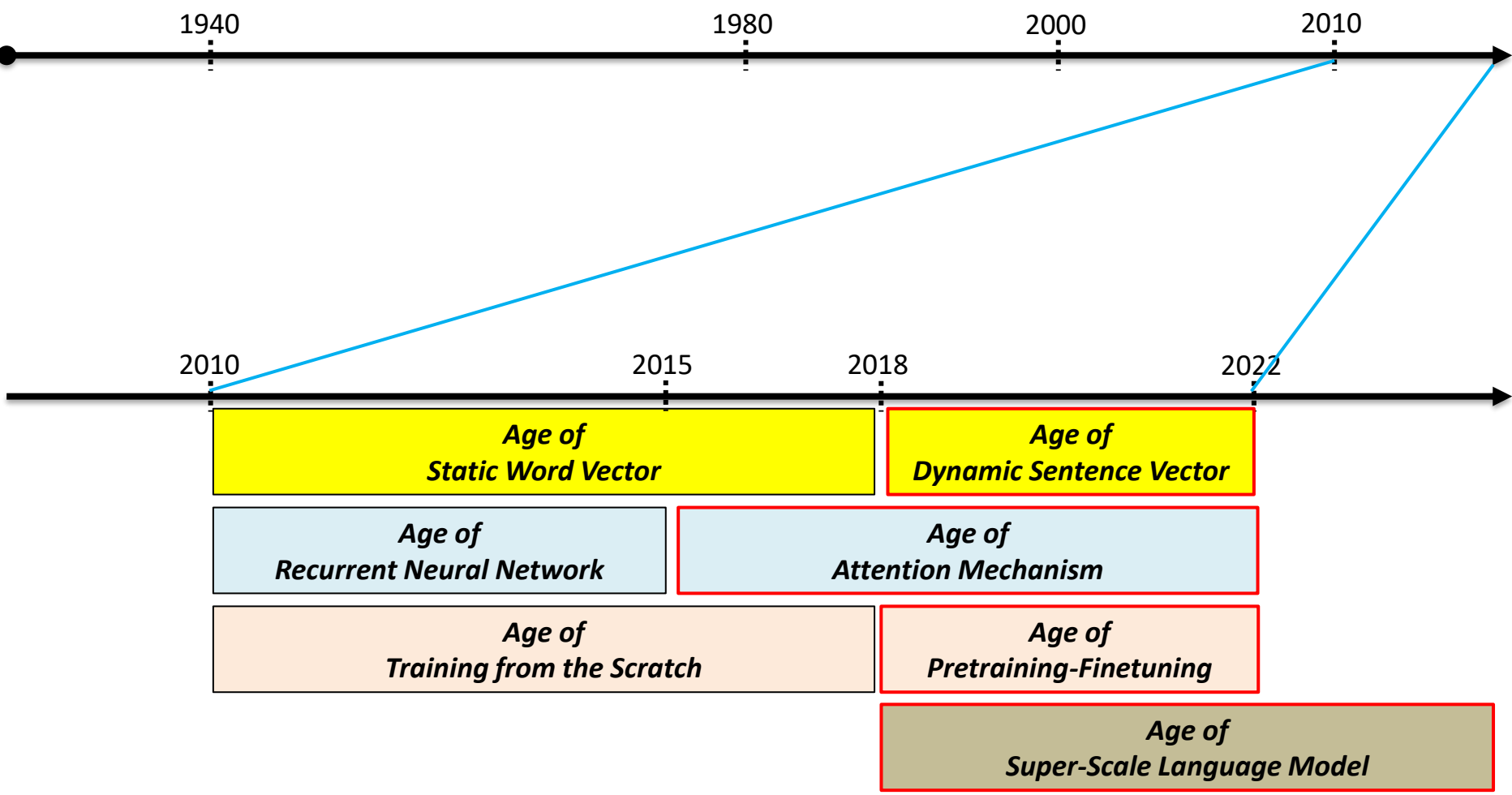


Throughout the generations,

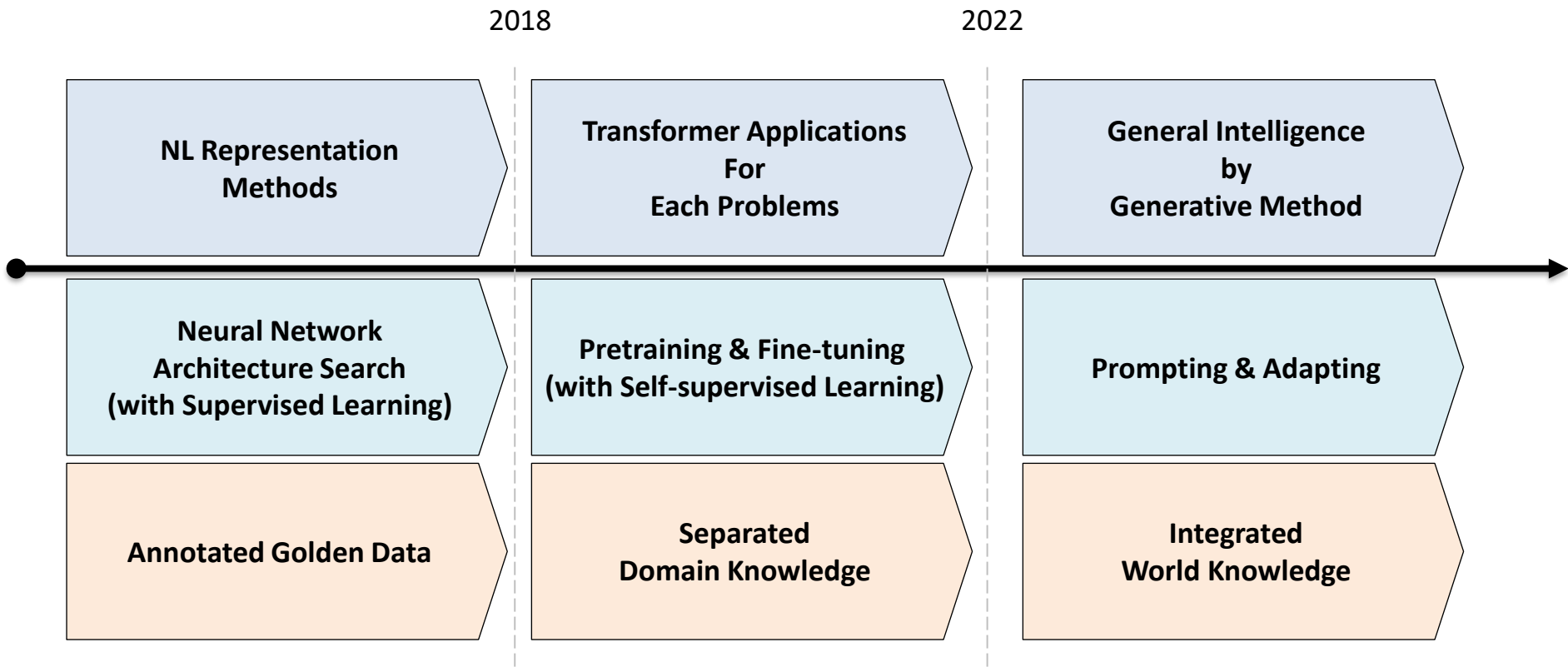
- The methods of Knowledge management,
- Natural language understanding formats,
- And decision-making processes

Have completely changed.

Historical View - Artificial Intelligence (NLP)



Mega-Trend



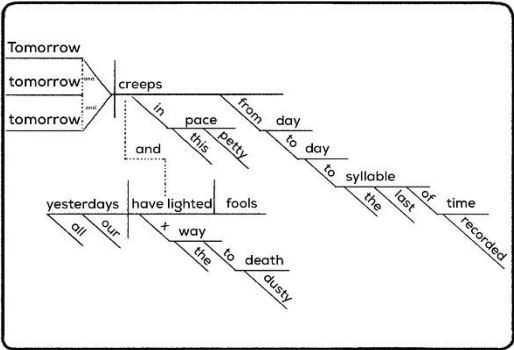
Problem Definition - History

1970

2022

General Language Rule

- ✓ Universal Grammar



<https://practicalpie.com/chomskys-universal-grammar/>

Information Extraction

- ✓ Extract (Structural) information form unstructured natural language
- ✓ Naturally, the classification technique is well-suited.

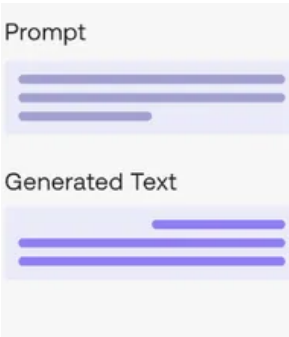
When **Sebastian Thrun** started at **Google** in **2007**, few people outside of the company took him seriously. "I can tell you very senior CEOs of major **American** car companies would shake my hand and turn away because I wasn't worth talking to," said **Thrun**, now the co-founder and CEO of online higher education startup Udacity, in an interview with **Recode** **earlier this week**.

A little **less than a decade later**, dozens of self-driving startups have cropped up while automakers around the world clamor, wallet in hand, to secure their place in the fast-moving world of fully automated transportation.

<https://www.aitimejournal.com/@akshay.chavan/complete-tutorial-on-named-entity-recognition-ner-using-python-and-keras>

Answer Generation

- ✓ Generate the most likable outputs that might follow the preceding text.



<https://www.analyticsvidhya.com/blog/2023/05/what-is-prompt-engineering-guide/>

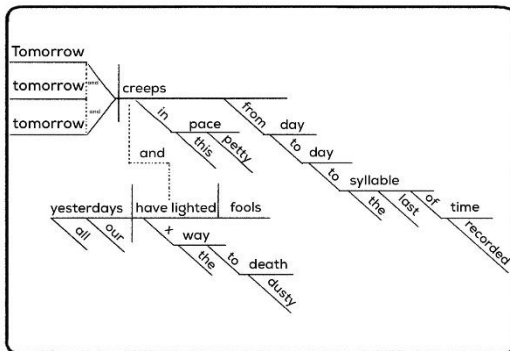
COURSE OVERVIEW

1970

Course Coverage

General Language Rule

- ✓ Universal Grammar



<https://practicalpie.com/chomskys-universal-grammar/>

Information Extraction

- ✓ Extract (Structural) information from unstructured natural language
- ✓ Naturally, the classification technique is well-suited.

When **Sebastian Thrun** started at **Google** in **2007**, few people outside of the company took him seriously. "I can tell you very senior CEOs of major **American** car companies would shake my hand and turn away because I wasn't worth talking to," said **Thrun**, now the co-founder and CEO of online higher education startup Udacity, in an interview with **Recode** **earlier this week**.

A little **less than a decade later**, dozens of self-driving startups have cropped up while automakers around the world clamor, wallet in hand, to secure their place in the fast-moving world of fully automated transportation.

<https://www.aitimejournal.com/@akshay.chavan/complete-tutorial-on-named-entity-recognition-ner-using-python-and-keras>

Answer Generation

- ✓ Generate the most likable outputs that might follow the preceding text.

Prompt

Generated Text

<https://www.analyticsvidhya.com/blog/2023/05/what-is-prompt-engineering-guide/>

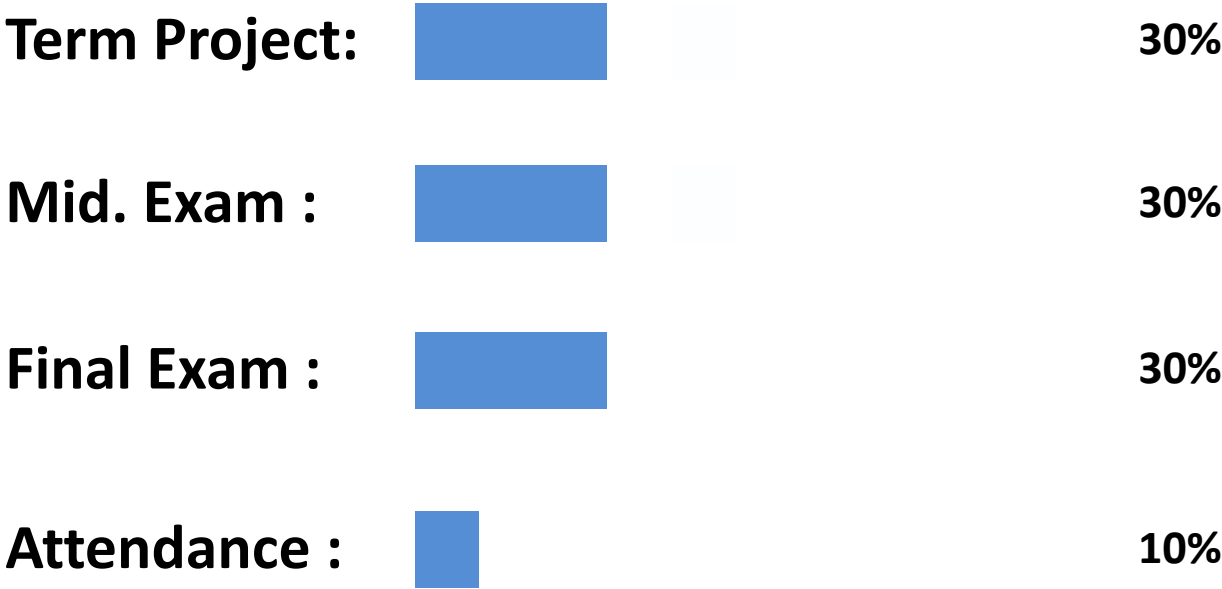
Course Overview

Neural Network based NLP	Transformer		Transformer NLPs
<ul style="list-style-type: none">✓ Tensor Operations✓ NN design pattern✓ PyTorch-Lightning✓ Seq2Seq✓ Attention	<ul style="list-style-type: none">✓ Transformer Overview✓ QKV pattern✓ Multihead Attention✓ Self- and Cross-Attention✓ Scaling, Normalization, Residual connection	<div>[Code Practice]</div> <ul style="list-style-type: none">✓ QKV Attention✓ Multi-head attention✓ Encoder✓ Encoder (Multi-layer)✓ Encoder only - BERT✓ Original Transformer (Sorting Problem)✓ Decoder only – GPT2	<ul style="list-style-type: none">✓ NLP Problems Overview✓ N21 problem✓ N2N problem✓ N2M problem
Evaluation Metrics	Generative LM		LLM-based Application Dev.
	<ul style="list-style-type: none">✓ Understanding Llama1, 2 and 3✓ Prompt✓ Prompt tuning✓ Adapting✓ ...		<ul style="list-style-type: none">✓ Term Projects

Course Overview

<i>Week 1</i>	Syllabus, NLP Overview
<i>Week 2</i>	Neural Network based NLP
<i>Week 3</i>	Transformer (1) Online Practice
<i>Week 4</i>	Transformer (2) Online Practice
<i>Week 5</i>	Hugging Face – Tool Introduction, NLP Performance Analysis
<i>Week 6</i>	How to Improve NLP applications, N21 Problem Practice
<i>Week 7</i>	N2N Problem, N2M Problem Practice
<i>Week 8</i>	Midterm Exam.
<i>Week 9</i>	Publish Model and Data Practice
<i>Week 10</i>	Term Project Proposal
<i>Week 11</i>	Llama (1)
<i>Week 12</i>	Llama (2)
<i>Week 13</i>	LLM Training Term Project Coaching
<i>Week 14</i>	Final Exam.
<i>Week 15</i>	Term Project Presentation

Evaluation



Contact Information | Professor



Sangkeun Jung

E-mail : hugmanskj@gmail.com,
hugman@cnu.ac.kr

Laboratory : Intelligent Software
Office : W5624