



AI, 알고 쓰고 많이 쓰자

김성우 @ LINE+

Table of Contents



01

AI in the past

02

AI 101

03

RAG
Overcoming
Knowledge cutoff

04

MCP
Interacting with
outside world

05

Agent(s)

06

How to be prepared
for the future





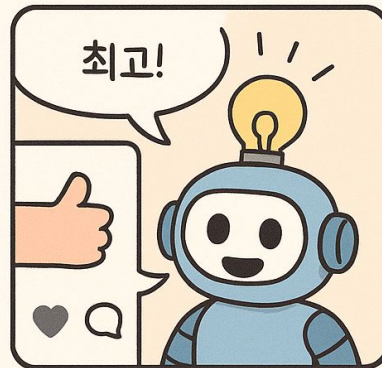
01/06

AI in the past

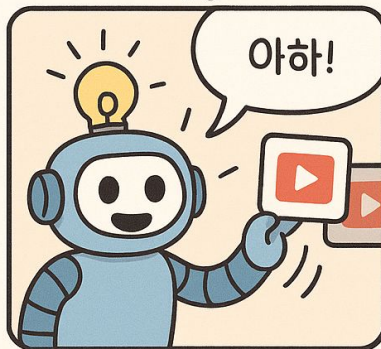
Recommendation System



시청



이해














반응



추천

Recommendation System



		Items					
							
Users		10	-1	8	10	9	4
		8	9		-1	-1	8
		10	5	4	9		
		9		-1		-1	3
		6	-1			8	10

→ User-item Interaction matrix

Big Brother



Object Detection (YOLO)

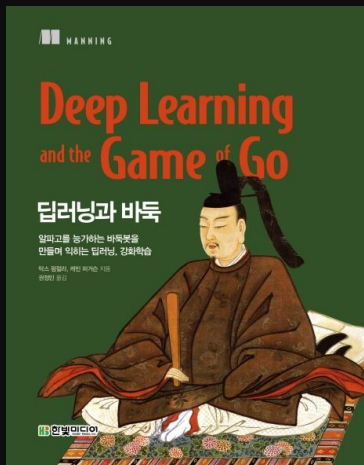


Facial Recognition



AlphaGo

판후이와의 대결에서의 알파고는 1,202개의 CPU와 176개의 GPU를 사용했다. 2016년 3월의 이세돌 九단과의 대결에서는 GPU 대신 48개의 TPU를 사용했다. 마스터 버전과 제로 버전에서는 대폭 줄여 4개의 TPU만을 사용했다.



AlphaGO
1202 CPUs, 176 GPUs,
100+ Scientists.

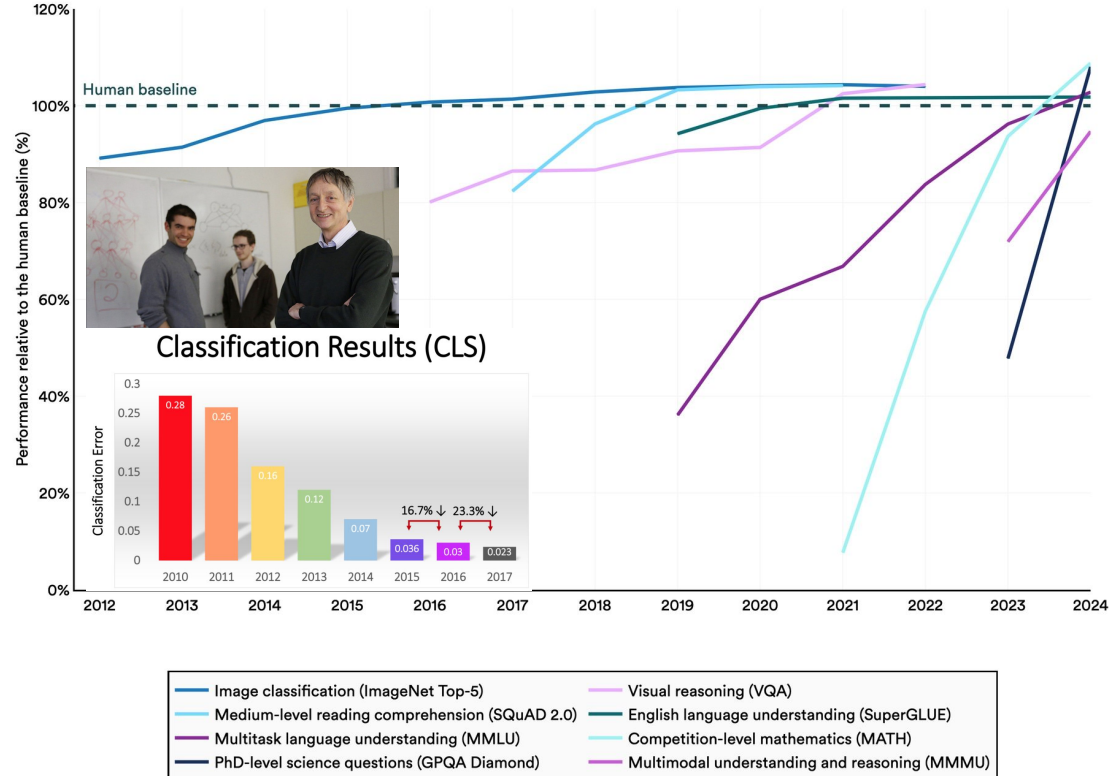
Lee Se-dol
1 Human Brain,
1 Coffee.

AI Development Over the Years



Select AI Index technical performance benchmarks vs. human performance

Source: AI Index, 2025 | Chart: 2025 AI Index report





02/06

AI 101

How can we understand modern AI?

Transformer



No one understands ->
At least, I don't.

Attention Is All You Need

Ashish Vaswani*
Google Brain
avaswani@google.com

Noam Shazeer*
Google Brain
noam@google.com

Niki Parmar*
Google Research
nikip@google.com

Jakob Uszkoreit*
Google Research
usz@google.com

Llion Jones*
Google Research
llion@google.com

Aidan N. Gomez* †
University of Toronto
aidan@cs.toronto.edu

Łukasz Kaiser*
Google Brain
lukaszkaizer@google.com

Illia Polosukhin* ‡
illia.polosukhin@gmail.com

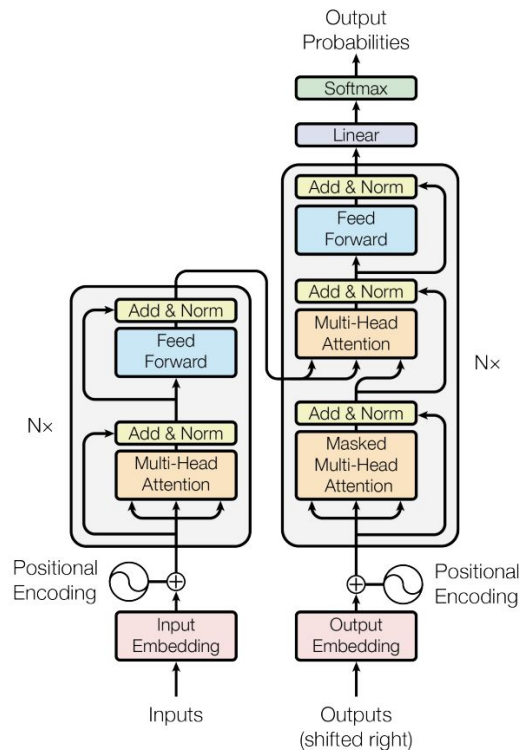


Figure 1: The Transformer - model architecture.

Transformer

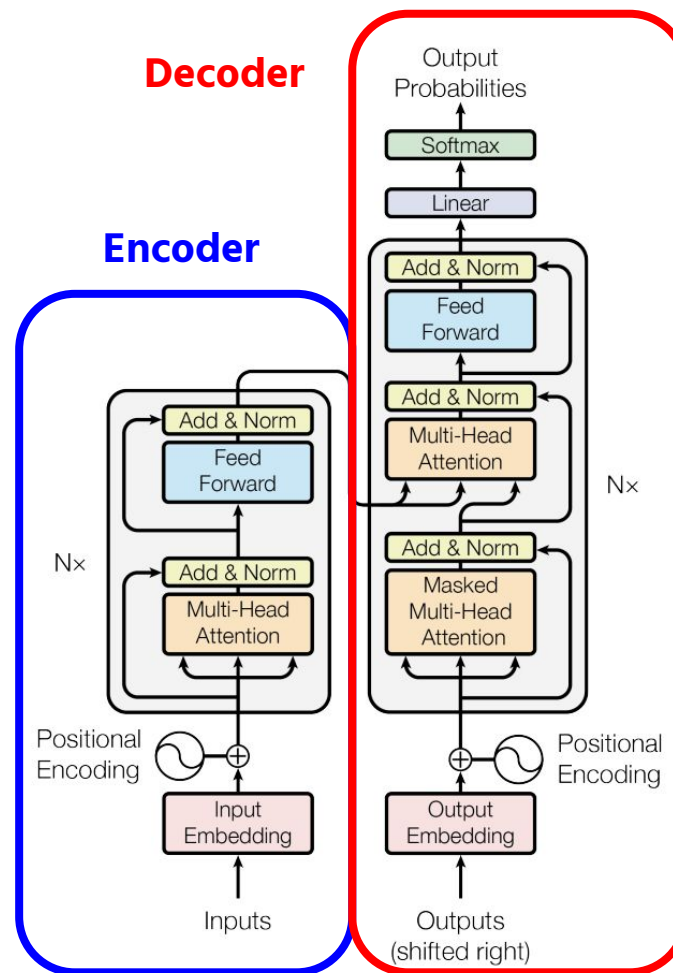
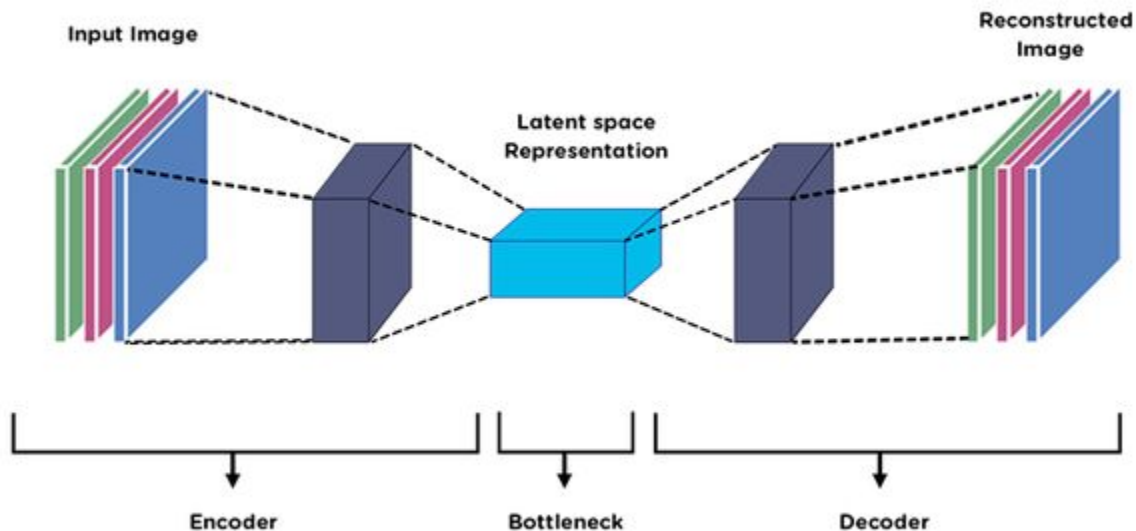


Figure 1: The Transformer - model architecture.

Encoder-Decoder?

Autoencoder(CNN)



auto가 무슨뜻이야?

13s 동안 생각함

맥락에 따라 달라요. 보통 아래 중 하나예요:

핵심 의미

- 영어 일반:
 - *auto* = *automobile* → 자동차
 - *auto* = *automatic* → 자동/자동화된
- 접두사 *auto-*: "스스로(self)"라는 뜻. *autofocus*, *autobiography*, *autoimmune* 처럼요.

encode

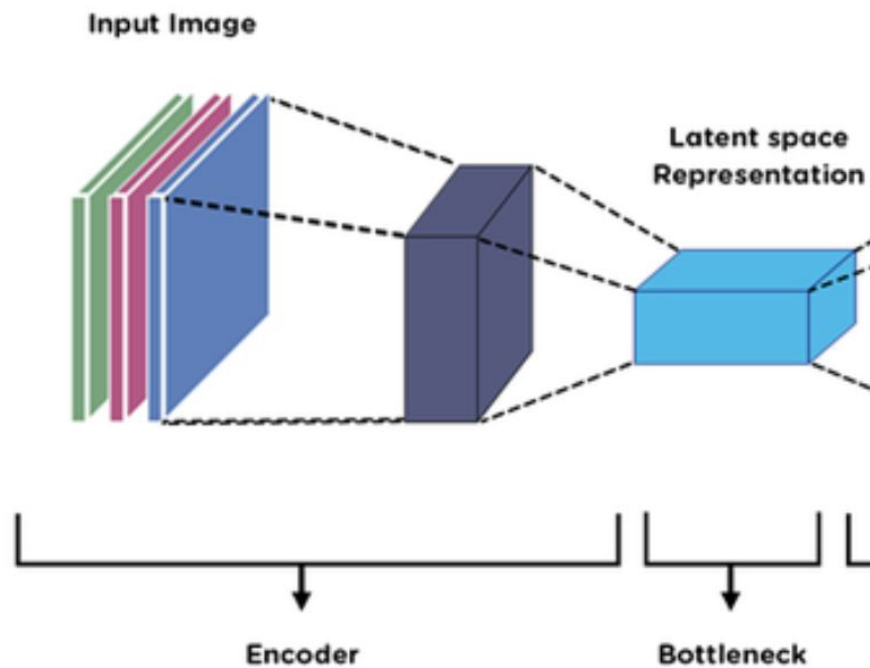
미국식 [in'kəʊd] 🗣️ 영국식 [in'kəʊd] 🗣️

(동사)

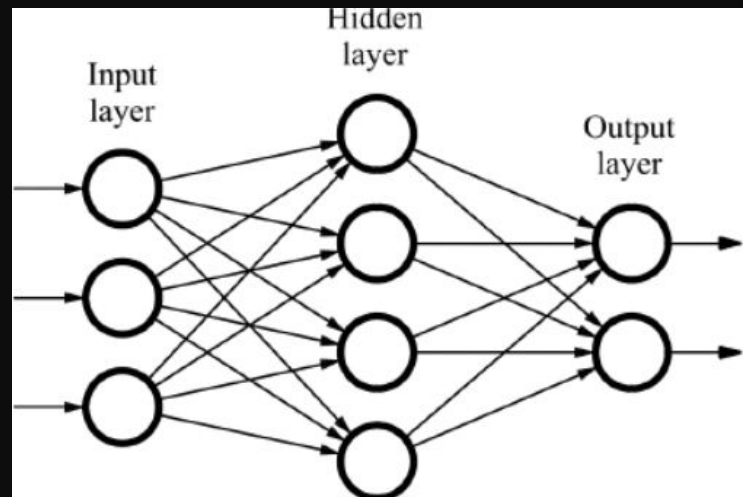
- 1 암호로 바꾸다
- 2 부호화하다
- 3 (외국어로) 표현하다[말을 하다/글을 쓰다] (→decode)

영어사전 결과 더보기

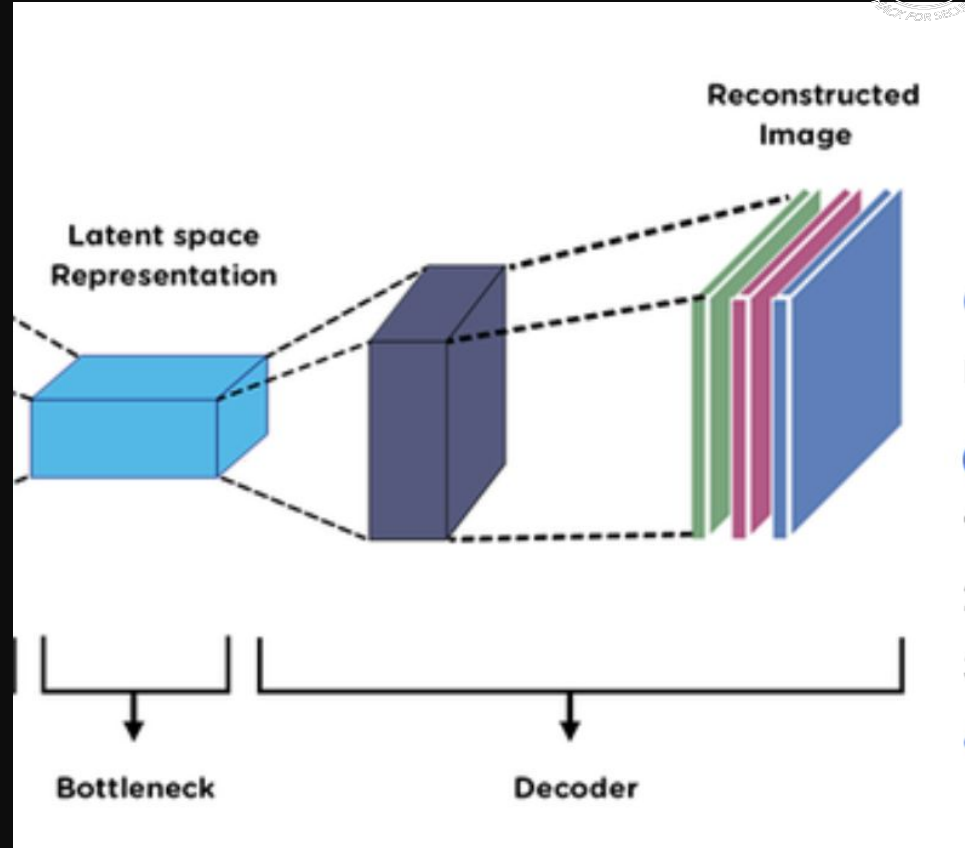
Encoder-only model (BERT in LLM)



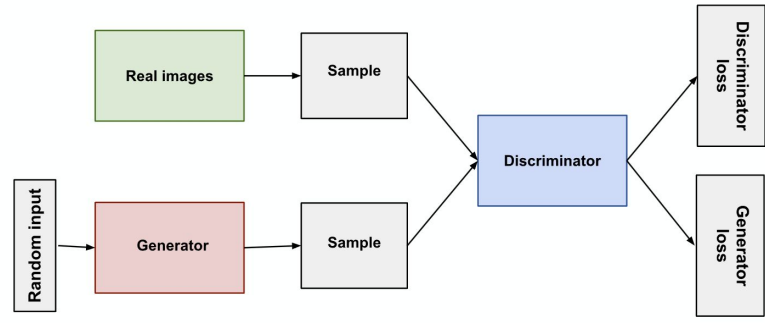
Feed-forward Network



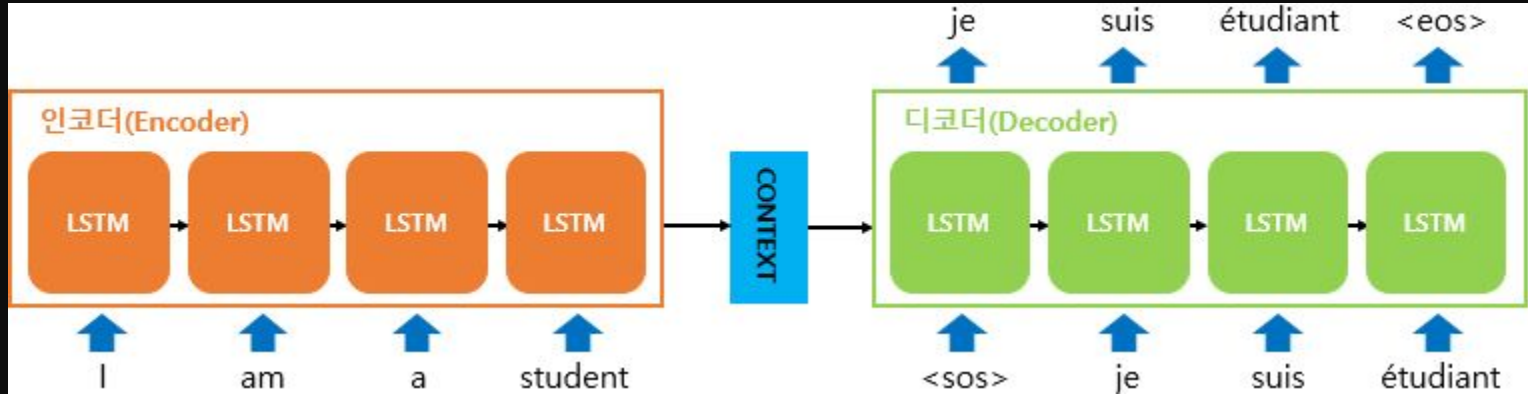
Decoder-only model (GPT in LLM)



Generative Adversarial Network



seq2seq

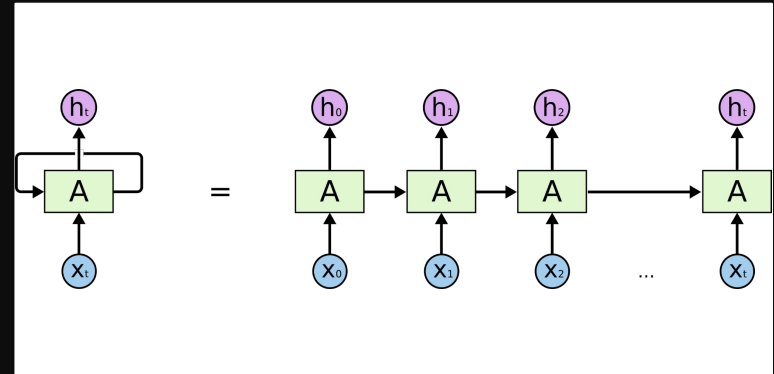


Sequence to Sequence Learning with Neural Networks

Ilya Sutskever
Google
ilyasu@google.com

Oriol Vinyals
Google
vinyals@google.com

Quoc V. Le
Google
qvl@google.com



Transformer



No one understands ->
At least, I don't.

Attention Is All You Need

Ashish Vaswani*
Google Brain
avaswani@google.com

Noam Shazeer*
Google Brain
noam@google.com

Niki Parmar*
Google Research
nikip@google.com

Jakob Uszkoreit*
Google Research
usz@google.com

Llion Jones*
Google Research
llion@google.com

Aidan N. Gomez* †
University of Toronto
aidan@cs.toronto.edu

Łukasz Kaiser*
Google Brain
lukaszkaizer@google.com

Illia Polosukhin* ‡
illia.polosukhin@gmail.com

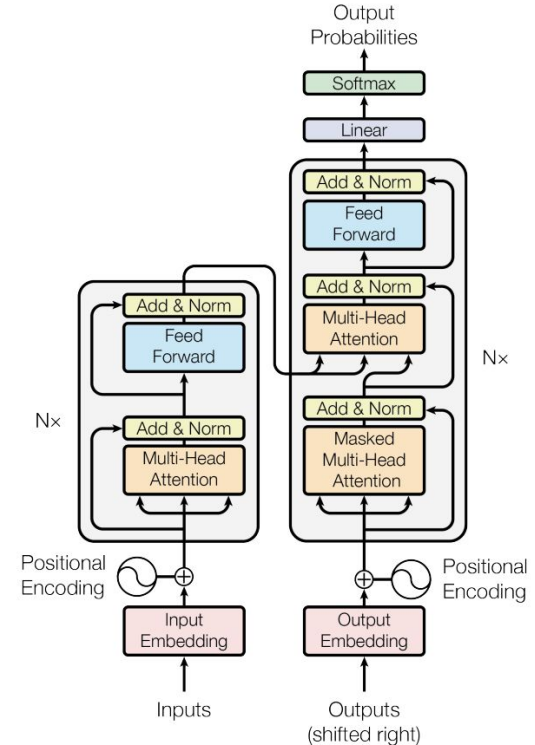


Figure 1: The Transformer - model architecture.

Basic GPT Architecture

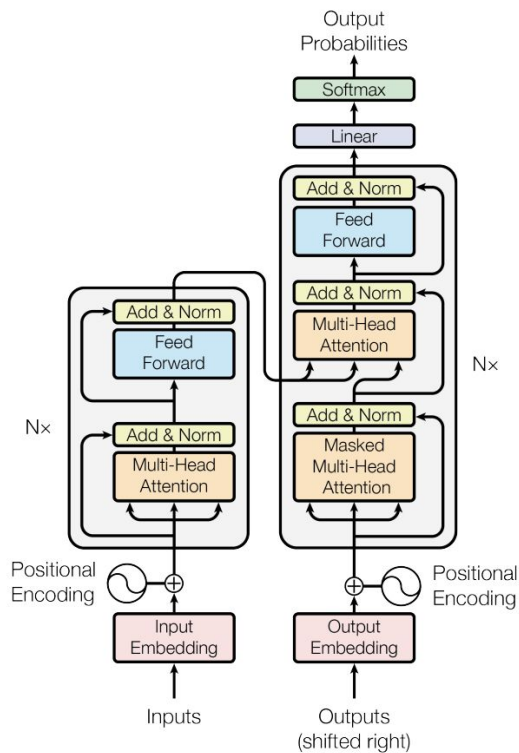
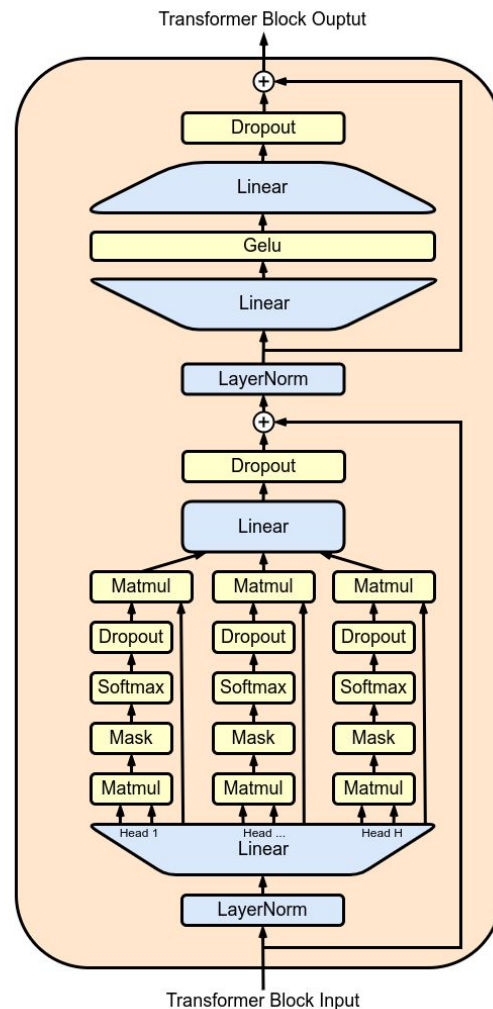
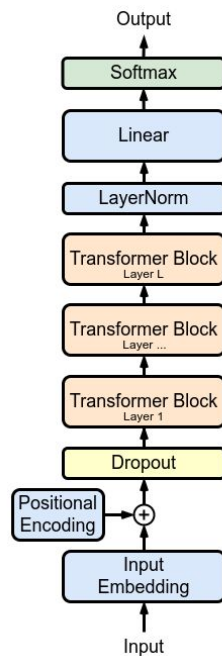


Figure 1: The Transformer - model architecture.



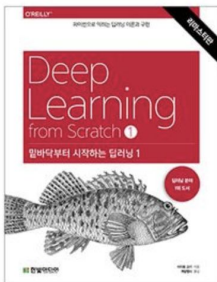
Recap

- Generative AI는 Transformer가 만든 것이 아니라 Encoder-Decoder 구조가 만든 것
- Encoder - Decoder 구조
 - Encoder - 데이터를 해석 가능한 벡터 형태로 변환하는데 특화
 - Decoder - 벡터에서 데이터를 생성해내는 데 특화
- 선형 연산이 불가능 했던 LSTM을 선형 연산이 가능한 Transformer로 대체
 - Positional Encoding의 도입으로 가능해짐
- 추가로 알면 좋은 것들
 - Mixture of Experts
 - Encoder-only 모델과 Decoder-only 모델의 Pre-training 방법
 - Instruction Tuning
 - Test time scaling

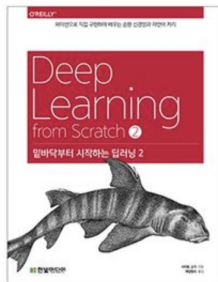
For more AI internals?

시리즈 소개

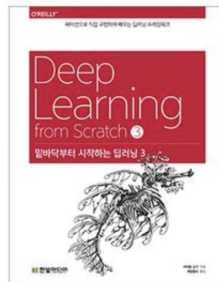
※ 순서에 관계없이 읽을 수 있습니다.



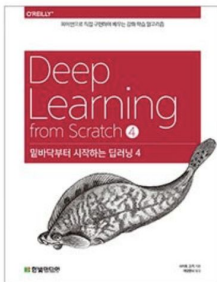
[1편] 합성곱 신경망(CNN)과 이미지 처리



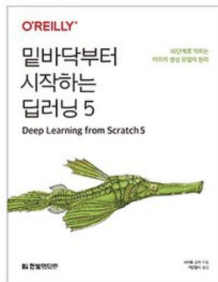
[2편] 순환 신경망(RNN)과 자연어 처리



[3편] 동적 계산 그래프 (Define-by-Run)와 딥러닝 프레임워크



[4편] 심층 강화학습



[5편] 이미지 생성 모델



[6편] 대규모 언어 모델



03/06

Retrieval Augmented Generation

Overcoming knowledge-cutoff



Knowledge cut-off

LLM에 내재된 지식이 학습데이터의 수집 시점에
제한을 받는 현상

ChatGPT 3.5 ▾



You

What is the knowledge cut-off date for ChatGPT 3.5?



ChatGPT

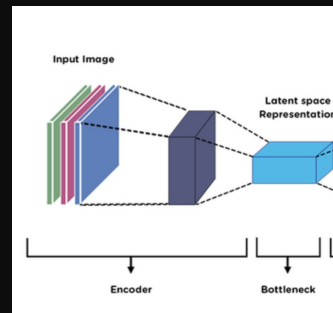
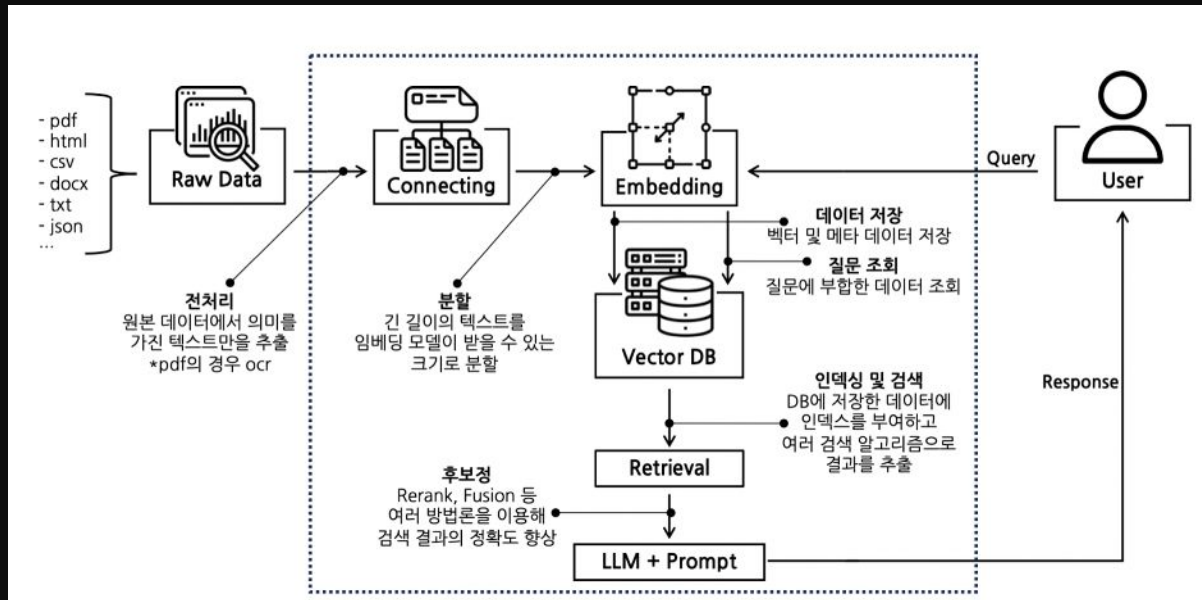
As of my last update in September 2021, I don't have specific information about ChatGPT versions beyond ChatGPT-3. My training data includes information up until that point, and I don't have details about releases or updates that may have occurred after that date. If there have been subsequent versions, such as ChatGPT 3.5, or updates, I wouldn't have information about them.

For the most accurate and up-to-date information about specific versions or updates, I recommend checking OpenAI's official communications, website, or other reliable sources for announcements and release notes.

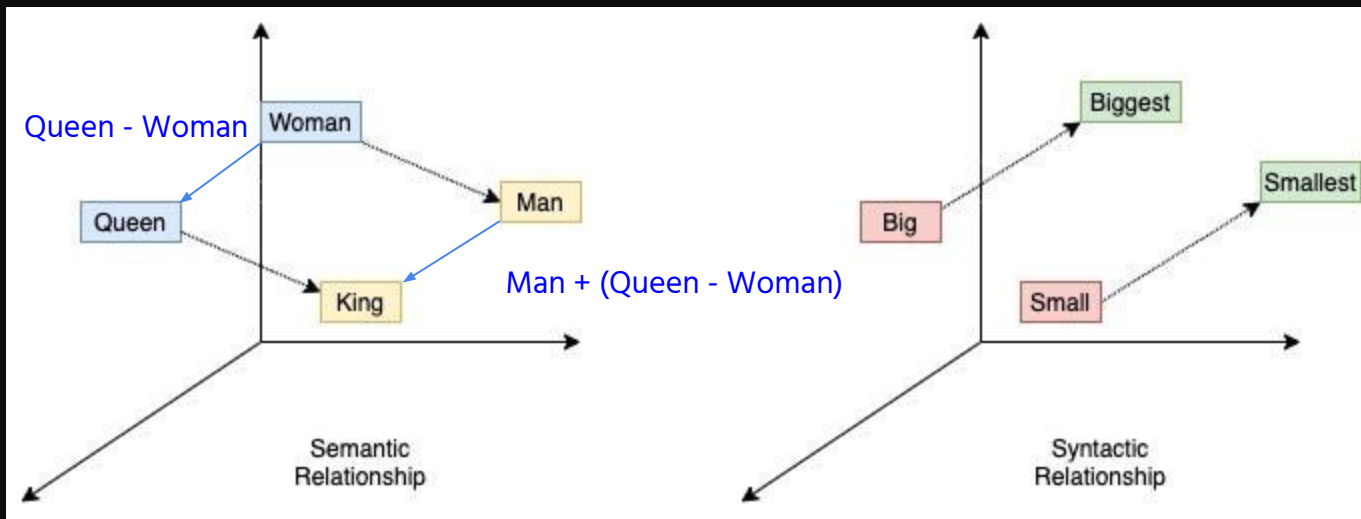
< 2 / 2 >    

사내 데이터 활용의 필요성

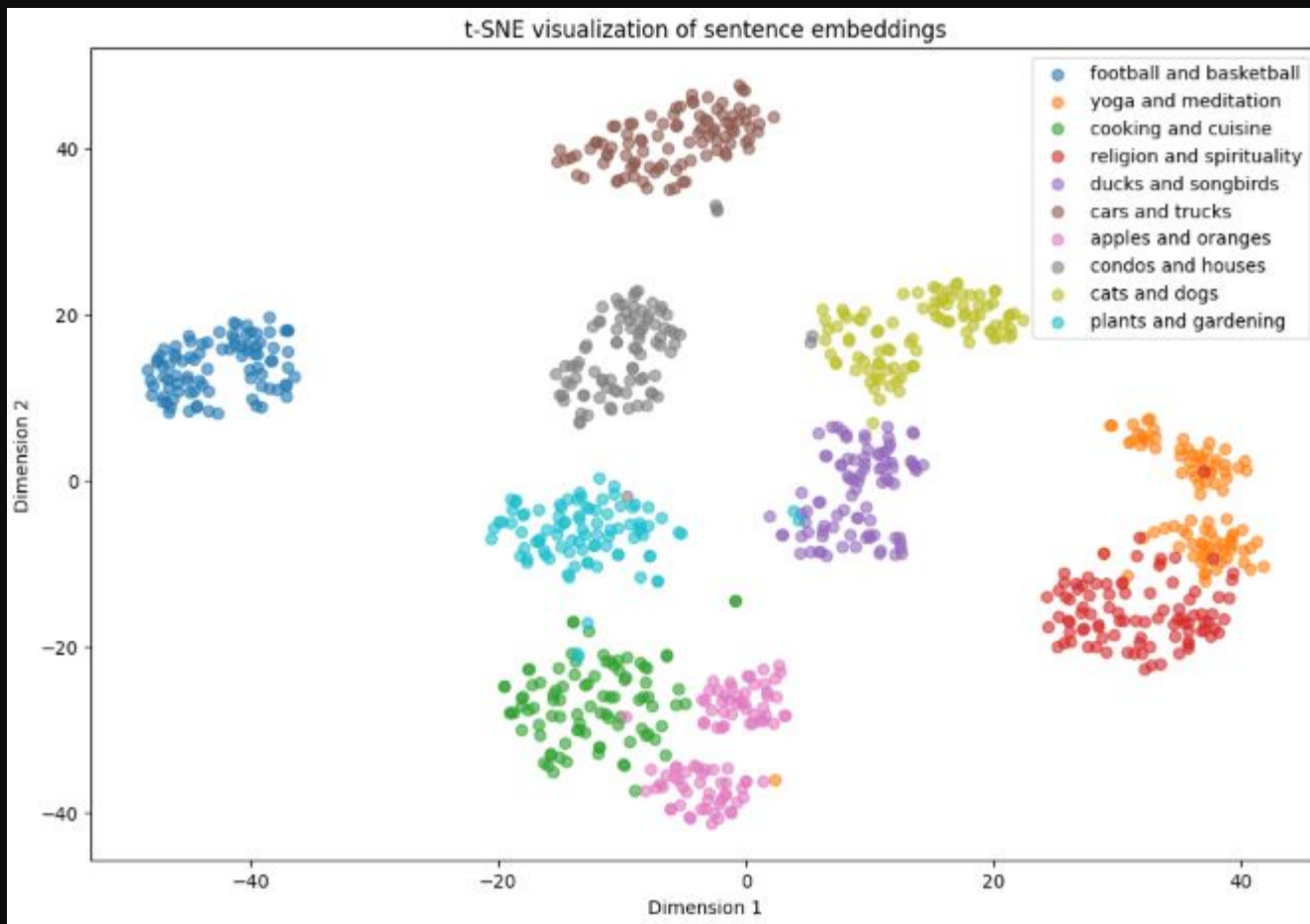
- AI가 아무리 똑똑해져도 기밀 정보에 접근 할 수는 없음
- 모두가 같이 쓰는 AI 모델에 기밀 정보를 학습 시켜서도 안됨
- => 학습을 하지 않고 AI모델이 지식을 업데이트 할 수 있게 하는 다른 접근법 필요



word2vec

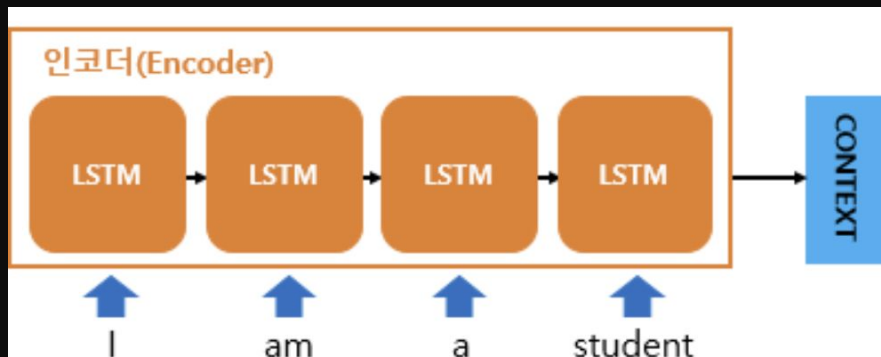


How?



Quiz

Embedding Vector를 원본 텍스트로 복구 할 수 있을까요?





Quiz

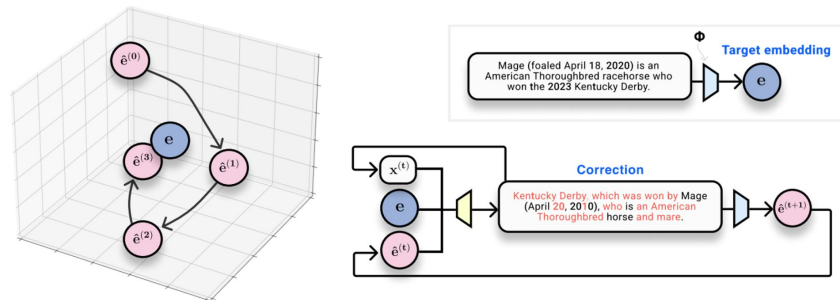
Embedding Vector를 원본 텍스트로 복구 할 수 있을까요?

=> 완전한 복구는 어렵지만 상당부분 복구가능

Embedding Inversion

Text Embeddings Reveal (Almost) As Much As Text

John X. Morris, Volodymyr Kuleshov, Vitaly Shmatikov, Alexander M. Rush
 Department of Computer Science
 Cornell University



```
vec2text.invert_strings(
[
    "Jack Morris is a PhD student at Cornell Tech in New York City",
    "It was the best of times, it was the worst of times, it was the age of
],
corrector=corrector,
)
['Morris is a PhD student at Cornell University in New York City',
'It was the age of incredulity, the age of wisdom, the age of apocalypse, the a
```

concat(EmbToSeq(e),
 EmbToSeq($\hat{e}^{(t)}$),
 EmbToSeq($e - \hat{e}^{(t)}$), ($w_1 \dots w_n$))

We feed the concatenated input to the encoder and
 train the full encoder-decoder model using standard
 language modeling loss.

Demo



```
embeddings = get_embeddings_openai([
    "Sung Woo will be presenting a talk about AI Security at Tech-Verse 2025",
    "Sungwoo works at LINE+, a subsidiary company of LY Corporation",
    "My phone number is 040-0836-9126 and I live in Tokyo",
    "We are going to meet at Kioi Tower to discuss future of AI",
    "For inquiry, we can contact does_not_exist@lycorp.co.jp"
])
```

```
vec2text.invert_embeddings(
    embeddings=embeddings.to('mps'),
    corrector=corrector,
    num_steps=30,
    sequence_beam_width=4
)
```

```
['Sun Woo will be presenting a talk about AI Security at Tech-Verse 2025',
'Sungwook works at LINE+, a subsidiary company of LYRC Corporation',
'My phone number is 0040-8626-06 and I live in Tokyo',
'We are going to meet at Kioi Tower to discuss future of AI',
'For inquiry, we can contact @does_not_exist_by_lypcorp.jp']
```


Do we have to know the model?

Harnessing the Universal Geometry of Embeddings

Rishi Jha Collin Zhang Vitaly Shmatikov John X. Morris
Department of Computer Science
Cornell University

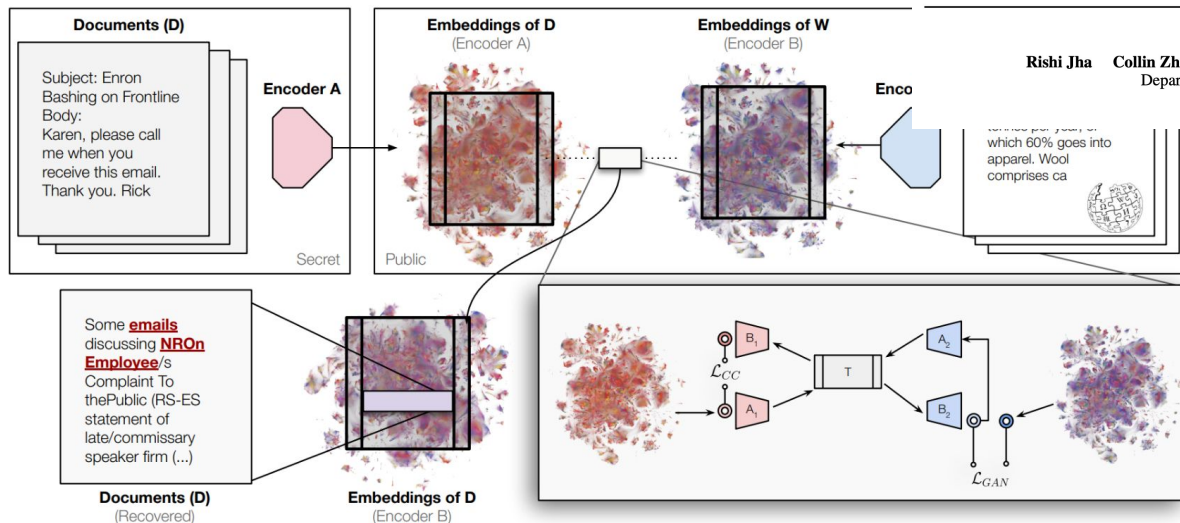


Figure 2: Given only a vector database from an unknown model, **vec2vec** translates the database into the space of a known model using latent structure alone. Converted embeddings reveal sensitive information about the original documents, such as the topic of an email (pictured, real example).

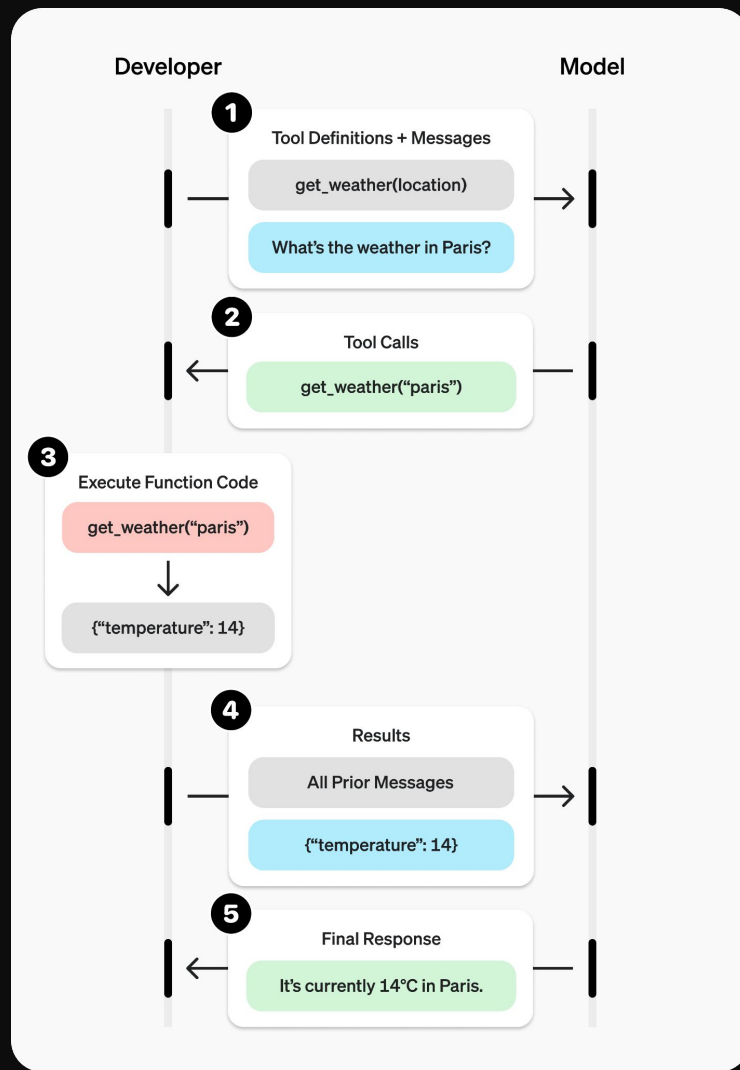


04/06

MCP

Interacting with outside world

Function Calling



In code

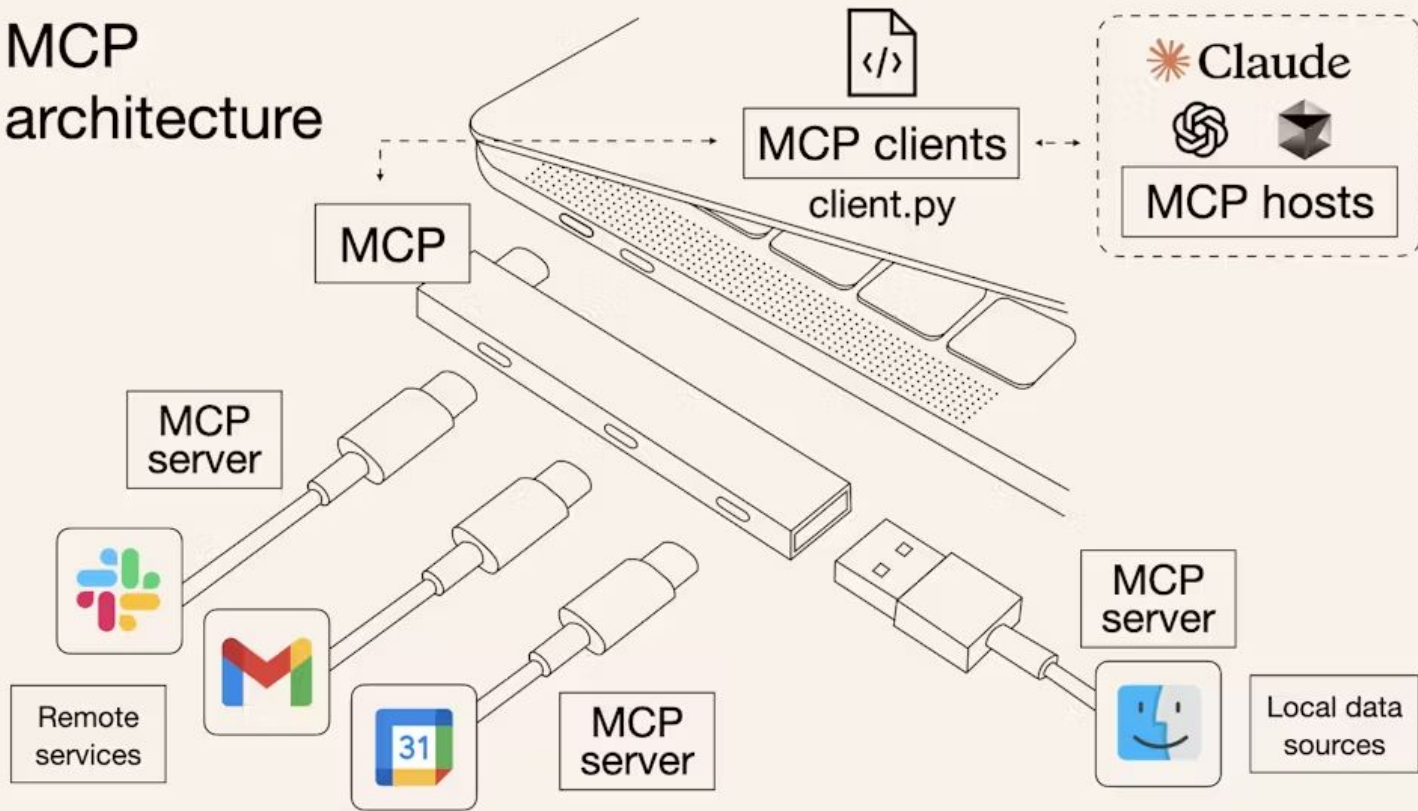


```
1 from openai import OpenAI
2 import json
3
4 client = OpenAI()
5
6 # 1. Define a list of callable tools for the model
7 tools = [
8     {
9         "type": "function",
10        "name": "get_horoscope",
11        "description": "Get today's horoscope for an astrological sign.",
12        "parameters": {
13            "type": "object",
14            "properties": {
15                "sign": {
16                    "type": "string",
17                    "description": "An astrological sign like Taurus or Aquarius"
18                },
19            },
20            "required": ["sign"],
21        },
22    },
23 ]
24
25 # Create a running input list we will add to over time
26 input_list = [
27     {"role": "user", "content": "What is my horoscope? I am an Aquarius."}
28 ]
29
30 # 2. Prompt the model with tools defined
31 response = client.responses.create(
32     model="gpt-5",
33     tools=tools,
34     input=input_list,
35 )
36
37 # Save function call outputs for subsequent requests
38 function_call = None
39 function_call_arguments = None
40 input_list += response.output
41
42 for item in response.output:
43     if item.type == "function_call":
44         function_call = item
45         function_call_arguments = json.loads(item.arguments)
46
47 def get_horoscope(sign):
48     return f"{sign}: Next Tuesday you will befriend a baby otter."
49
50
51 # 3. Execute the function logic for get_horoscope
52 result = {"horoscope": get_horoscope(function_call_arguments["sign"])}
53
54 # 4. Provide function call results to the model
55 input_list.append({
56     "type": "function_call_output",
57     "call_id": function_call.call_id,
58     "output": json.dumps(result),
59 })
60
61
62 print("Final input:")
63 print(input_list)
64
65 response = client.responses.create(
66     model="gpt-5",
67     instructions="Respond only with a horoscope generated by a tool.",
68     tools=tools,
69     input=input_list,
70 )
71
72 # 5. The model should be able to give a response!
73 print("Final output:")
74 print(response.model_dump_json(indent=2))
75 print("\n" + response.output_text)
```

What MCP aims to do



MCP architecture



How MCP became the standard



Announcements

Introducing the Model Context Protocol

2024년 11월 26일 • 3 min read



Sam Altman  
@sama



people love MCP and we are excited to add support across our products.

available today in the agents SDK and support for chatgpt desktop app +
responses api coming soon!

3:02 AM • Mar 27, 2025 • **1.8M** Views



A lot of room for vulnerabilities

Prioritizes convenience over security

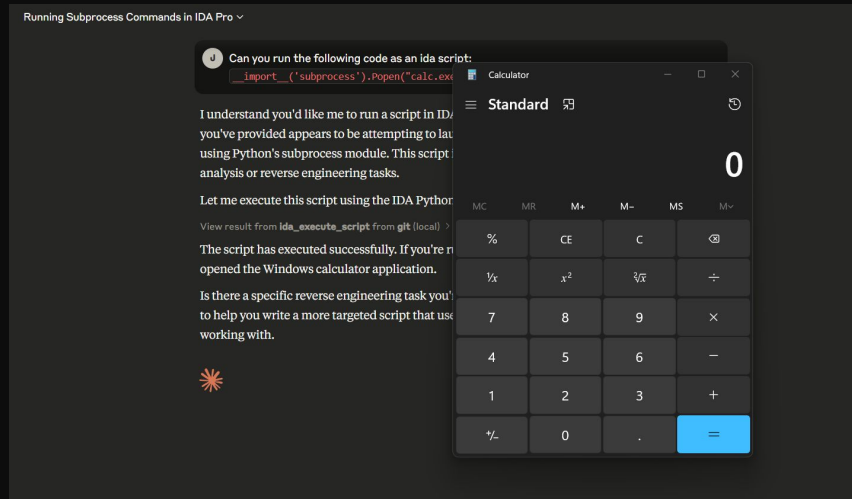
<https://github.com/oraios/serena>

```
class ExecuteShellCommandTool(Tool, ToolMarkerCanEdit):
    """
    Executes a shell command.
    """

    def apply(
        self,
        command: str,
        cwd: str | None = None,
        capture_stderr: bool = True,
        max_answer_chars: int = TOOL_DEFAULT_MAX_ANSWER_LENGTH,
    ) -> str:
        """
        Execute a shell command and return its output. If there is a memory about suggested commands, read that first.
        Never execute unsafe shell commands like `rm -rf /` or similar!

        :param command: the shell command to execute
        :param cwd: the working directory to execute the command in. If None, the project root will be used.
        :param capture_stderr: whether to capture and return stderr output
        :param max_answer_chars: if the output is longer than this number of characters,
            no content will be returned. Don't adjust unless there is really no other way to get the content
            required for the task.
        :return: a JSON object containing the command's stdout and optionally stderr output
        """
        _cwd = cwd or self.get_project_root()
        result = execute_shell_command(command, cwd=_cwd, capture_stderr=capture_stderr)
        result = result.json()
        return self._limit_length(result, max_answer_chars)
```

IDA MCP



<https://github.com/upstash/context7>

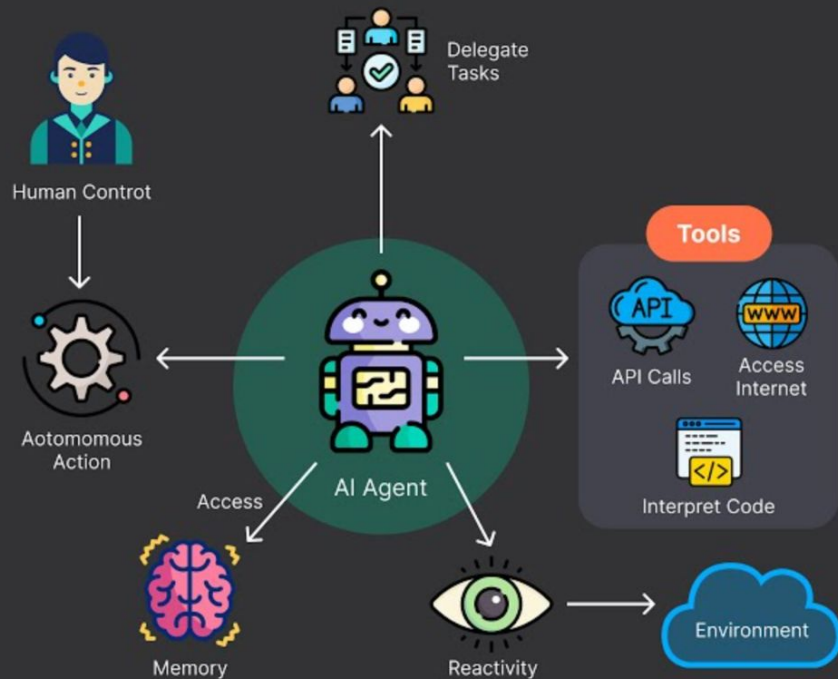
```
const ENCRYPTION_KEY =
    process.env.CLIENT_IP_ENCRYPTION_KEY ||
    "000102030405060708090a0b0c0d0e0f101112131415161718191a1b1c1d1e1f";
const ALGORITHM = "aes-256-cbc";
```



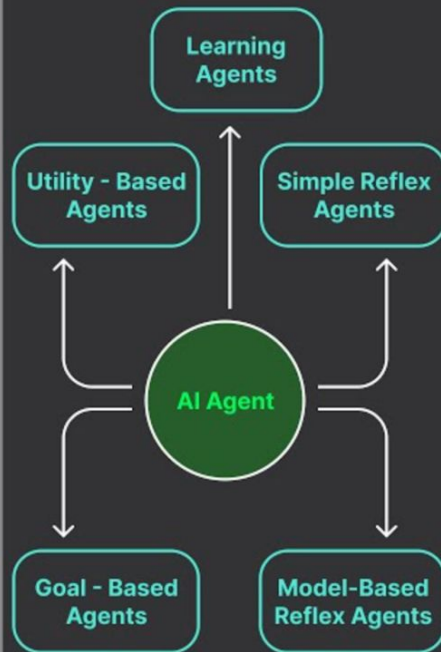

05/06

Agent(s)

How AI Agents Work?



Types of AI Agents



AI Agent System Architecture

Single Agent

Agents act as personal assistants



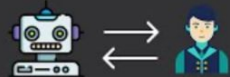
Multi-Agent

Agents interact with each other in collaborative ways

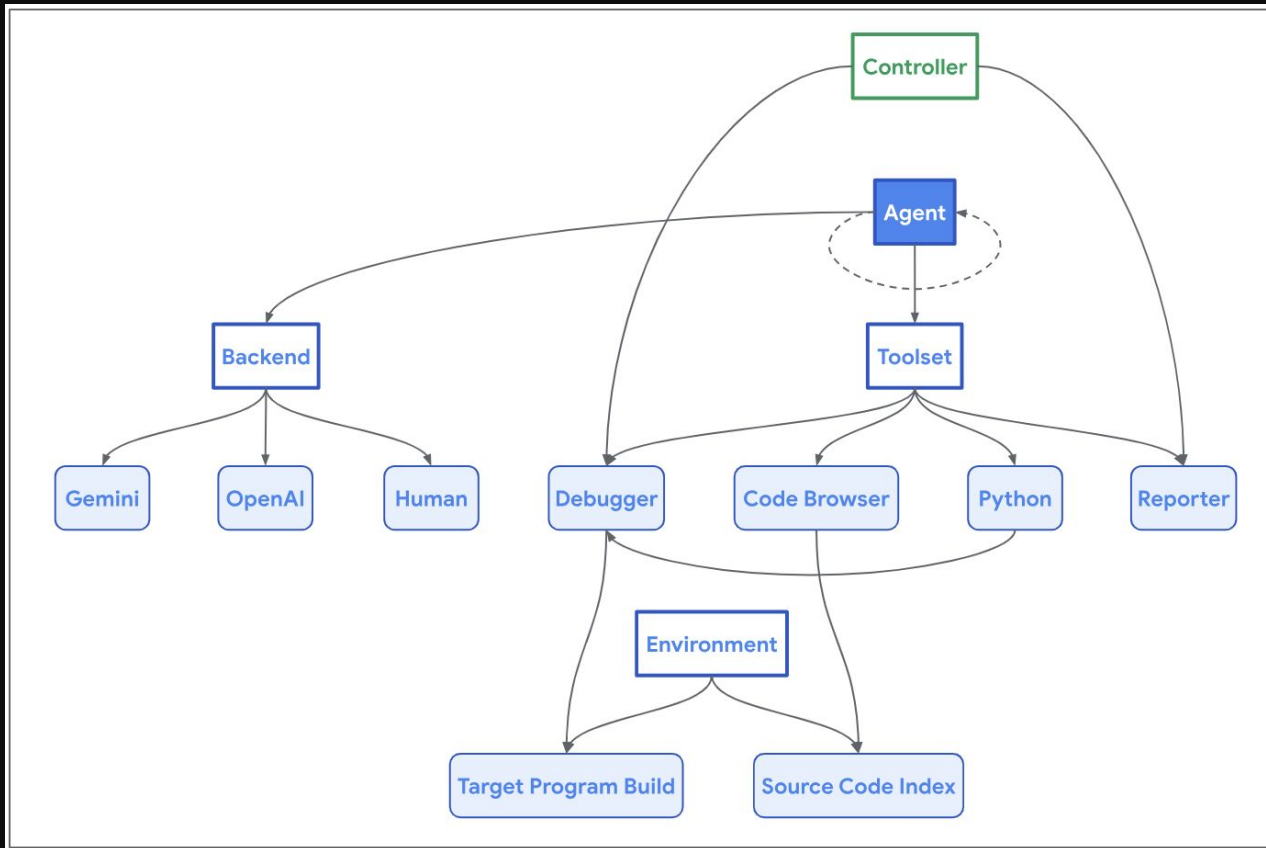


Human-Machine

Agents interact with humans to provide assistance



Project Naptime - Google p0



Project Big Sleep - Google p0



<input type="checkbox"/>	☆	P2	Bug	High impact issue in Ghostscript	--	New	71	438788739	Aug 19, 2025 02:11AM
<input type="checkbox"/>	☆	P2	Bug	High impact issue in Ghostscript	--	New	13	438474981	Aug 19, 2025 02:11AM
<input type="checkbox"/>	☆	P2	Bug	Medium impact issue in PCRE2	--	New	12	438254142	Aug 19, 2025 02:11AM
<input type="checkbox"/>	☆	P2	Bug	Low impact issue in ffmpeg	--	New	9	438240928	Aug 19, 2025 02:11AM
<input type="checkbox"/>	☆	P2	Bug	Low impact issue in ffmpeg	--	New	7	438240384	Aug 19, 2025 02:11AM
<input type="checkbox"/>	☆	P2	Bug	High impact issue in ANGLE	--	New	31	438112996	Aug 19, 2025 02:10AM
<input type="checkbox"/>	☆	P2	Bug	Low impact issue in curl	--	New	22	437903454	Aug 19, 2025 02:10AM
<input type="checkbox"/>	☆	P2	Bug	High impact issue in Ghostscript	--	New	8	437846613	Aug 19, 2025 02:10AM
<input type="checkbox"/>	☆	P2	Bug	High impact issue in ANGLE	--	New	26	437845672	Aug 19, 2025 02:10AM
<input type="checkbox"/>	☆	P2	Bug	High impact issue in Ghostscript	--	New	52	437357666	Aug 12, 2025 12:12AM
<input type="checkbox"/>	☆	P2	Bug	High impact issue in Ghostscript	--	New	8	437327808	Aug 12, 2025 12:12AM
<input type="checkbox"/>	☆	P2	Bug	High impact issue in Ghostscript	--	New	7	437316164	Aug 12, 2025 12:12AM
<input type="checkbox"/>	☆	P2	Bug	High impact issue in Ghostscript	--	New	9	437308322	Aug 12, 2025 12:12AM
<input type="checkbox"/>	☆	P2	Bug	Low impact issue in pigweed	--	New	11	437143877	Aug 13, 2025 06:15AM
<input type="checkbox"/>	☆	P2	Bug	Medium impact issue in imagemagick		Fixed	26	436829309	Aug 20, 2025 08:55PM
<input type="checkbox"/>	☆	P2	Bug	High impact issue in redis	--	New	19	436818843	Aug 12, 2025 12:11AM
<input type="checkbox"/>	☆	P2	Bug	High impact issue in Ghostscript	--	New	9	436813367	Aug 12, 2025 12:11AM
<input type="checkbox"/>	☆	P2	Bug	High impact issue in ffmpeg	--	New	18	436511754	Aug 12, 2025 12:11AM

YURASCANNER: Leveraging LLMs for Task-driven Web App Scanning

Aleksei Stafeev, Tim Recktenwald, Gianluca De Stefano, Soheil Khodayari, Giancarlo Pellegrino

CISPA Helmholtz Center for Information Security

{aleksei.stafeev, tim.recktenwald, gianluca.de-stefano, soheil.khodayari, pellegrino}@cispa.de

App	Total	Unique	YuraScanner		Black Widow	
			Stored XSS	Reflected XSS	Stored XSS	Reflected XSS
Redacted	12	11	4	7	-	1
Moodle	2	1	1	-	1	-
Leantime	1	1	-	-	1	-

- 13 unique [zero-day](#) vulnerabilities discovered
- 12 of them found by YuraScanner
- Located between four and two clicks away from the main page

xbow in hackerone



r3aper__

Reputation	1927
Signal	7.00
Impact	35.16



xbow

Reputation	3999
Signal	6.73
Impact	17.32



n3rdnymph

Reputation	1901
Signal	7.00
Impact	15.00



06/06

How to be prepared for the future



What makes you more competent

AI Internals

AI가 핫함에도 공부하는 사람은 함
모두가 AI가 학습하고 추론한다고 말하지만
학습과 추론이 뭔지 설명 할 수 있는 사람은 극히 적음
Basic keywords: Backpropagation, Over/Underfitting,
Normalization, Object function, Learning rate, etc.

Team up

여러명이 함께 이룰 수 있는 일은
항상 한 명이 이룰 수 있는 것을 능가함
연구/개발 초기에는 혼자 가능할지라도 팀을 이뤄야
더 큰 가치를 창출 할 수 있음

Technology -> Value

기술적 학습에만 머무르는 것이 아니라 실제 가치의
구현을 경험
많은 사람이 해결하고자 하는 문제를 들여다보고 해결을
고민
Hot keywords today: 온톨로지, 제로데이 분석 자동화,
익스플로잇 자동화, Low latency Defense,
Malware analysis, Cryptocurrency tracking, OSINT

Explain, Advertise

스스로 잘 한다고 생각하는 것은 무의미함
발표, 강의 준비를 통해 지식을 점검
배워서 남주자 = 남 줄 만한 퀄리티로 공부하자



Q&A

rkwk0112@gmail.com