

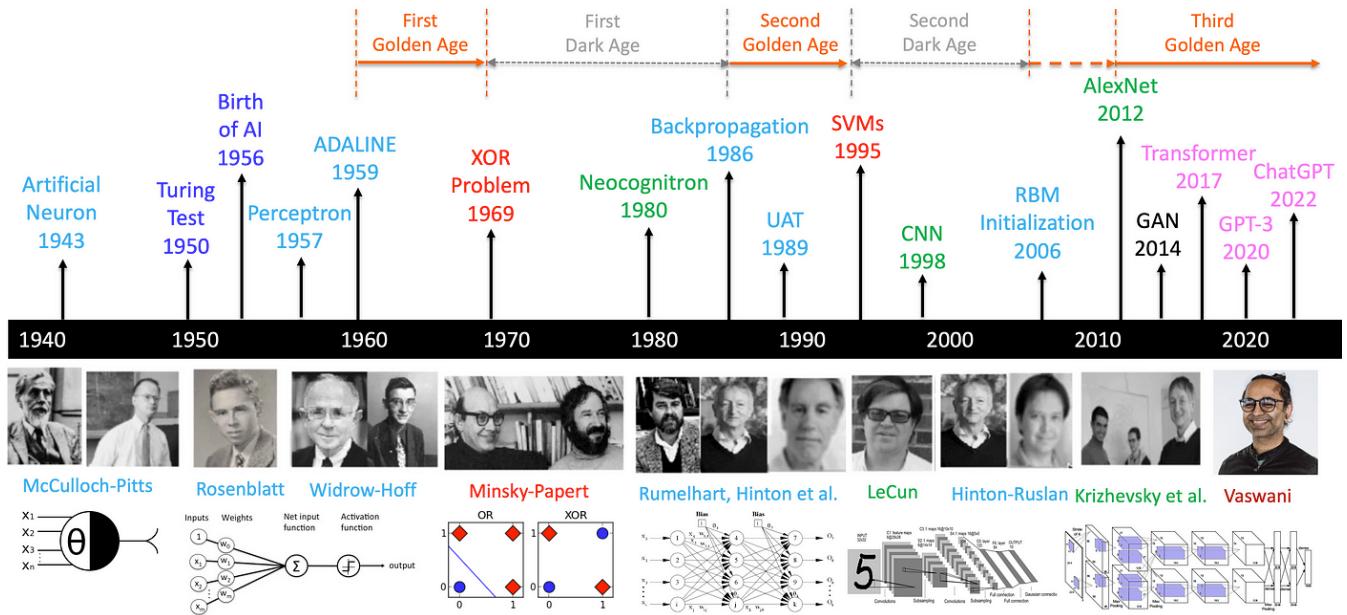
# Deep Learning 위주의 인공지능의 역사

📎 아래 링크의 원문을 DeepL로 번역하고 수정했습니다.

<https://medium.com/@lmpo/a-brief-history-of-ai-with-deep-learning-26f7948bc87b>

인공 지능(AI)과 딥 러닝은 지난 수십 년 동안 괄목할 만한 발전을 거듭하며 컴퓨터 비전, 자연어 처리, 로봇 공학 등의 분야를 변화시켰습니다. 이 문서에서는 초기 신경망 모델부터 최신 대규모 언어 모델과 멀티모달 AI 시스템에 이르기까지 딥러닝을 사용한 AI 역사의 주요 이정표를 간략하게 살펴봅니다.

## A Brief History of AI with Deep Learning



## 1. 인공 지능의 탄생 (1956)

인공 지능(AI)의 개념은 수세기 전부터 존재해 왔지만 오늘날 우리가 알고 있는 현대의 AI 분야는 20세기 중반부터 구체화되기 시작했습니다. '인공 지능'이라는 용어는 1956년 컴퓨터 과학자이자 인지 과학자인 존 매카시가 다트머스 하계 인공 지능 연구 프로젝트에서 처음 사용했습니다.

다트머스 컨퍼런스는 종종 연구 분야로서 인공지능의 발상지로 여겨집니다. 이 컨퍼런스에서는 컴퓨터 과학자, 수학자, 인지 과학자 그룹이 모여 인간의 지능을 시뮬레이션할 수 있는 기계를 만들 수 있는 가능성에 대해 논의했습니다. 참석자 중에는 마빈 민斯基, 나다니엘 로체스터, 클로드 셜넌과 같은 저명한 인사들이 포함되어 있었습니다.



다트머스 인공지능 여름 연구 프로젝트의 마빈 민스키, 클로드 쇄넌, 레이 솔로모노프 및 기타 과학자 (사진: 마가렛 민스키) [출처](#)

## 1.1 AI의 진화: 규칙 기반 시스템에서 딥 러닝까지

AI의 진화는 1950년대에 체스나 문제 해결과 같은 작업을 위한 알고리즘이 개발되면서 시작되었으며, 1956년 최초의 AI 프로그램인 논리적 이론가(Logical Theorist)가 탄생했습니다. 1960년대와 1970년대에는 복잡한 의사 결정 과정을 지원할 수 있는 MYCIN과 같은 규칙 기반 전문가 시스템이 도입되었습니다. 1980년대에는 AI 시스템이 데이터를 통해 학습하고 시간이 지남에 따라 개선할 수 있는 머신러닝이 등장하여 최신 딥러닝 기술의 토대가 마련되었습니다.

**DL ⊂ NN ⊂ ML ⊂ AI**

### Artificial Intelligence (AI)

- AI refers to a set of techniques that enable computers to mimic human behavior.

### Machine Learning (ML)

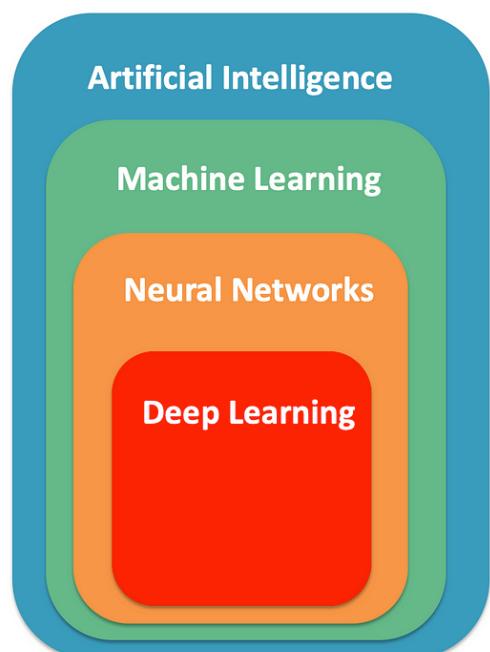
- ML is a subfield of AI, which enables machines to learn and make predictions based on data.

### Neural Networks (NN)

- NN are a subfield of ML that use artificial neural networks to extract patterns from data, inspired by the human brain.

### Deep Learning (DL)

- DL is a subfield of NN that utilizes multi-layered neural networks to achieve high performance on complex tasks.



오늘날 대부분의 최첨단 AI 기술은 딥러닝 기술을 기반으로 하고 있으며, 이는 AI의 지형을 변화시켰습니다. 머신러닝의 전문 분야인 딥러닝은 여러 계층으로 구성된 인공 신경망을 활용하여 원시 입력 데이터에서 복잡한 특징을 추출합니다. 이 글에서는 AI의 역사를 살펴보고 그 진화에서 딥러닝의 역할을 강조합니다.

## 2. 초기 인공 신경망(1940년대 - 1960년대)

### 2.1 뮤클로흐-피츠 뉴런(1943)

신경망의 개념은 1943년 웨伦 맥컬로크와 윌터 피츠가 최초의 인공 뉴런 모델을 제안했을 때로 거슬러 올라갑니다. 맥컬록-피츠 (MP) 뉴런 모델은 생물학적 뉴런을 획기적으로 단순화한 모델입니다. 이 모델은 이진 입력을 집계하고 임계값 활성화 함수를 사용하여 이 집계에 따라 결정을 내리고 이진 출력 {0, 1}을 산출함으로써 인공 신경망의 기초를 마련했습니다.

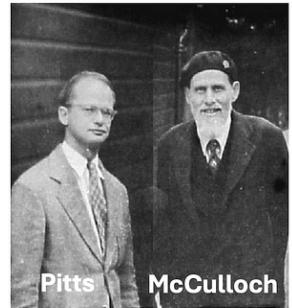
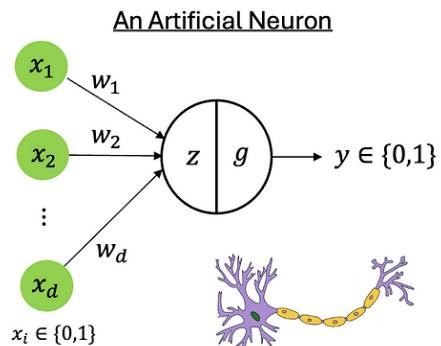
Both inputs and output are binary

- $x_i \in \{0,1\}$  and  $y \in \{0,1\}$

$$z(x_1, x_2, \dots, x_d) = z(\mathbf{x}) = \sum_{j=1}^d w_j \cdot x_j$$

$$y = g(z(\mathbf{x})) = \begin{cases} 1 & \text{if } z(\mathbf{x}) \geq T \\ 0 & \text{if } z(\mathbf{x}) < T \end{cases}$$

Threshold Activation function



이 단순화된 모델은 뉴런 행동의 본질을 포착합니다.

여러 입력을 수신하고, 이를 통합하고, 통합 신호가 임계값을 초과하는지 여부에 따라 이진 임계값을 초과하는지 여부에 따라 이진 출력을 생성합니다.

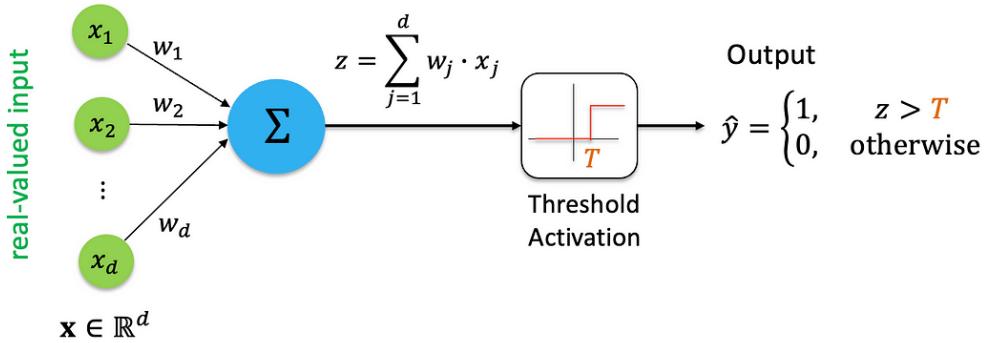
MP 뉴런 모델은 단순함에도 불구하고 다음과 같은 기능을 수행할 수 있습니다.

기본적인 논리 연산을 구현하여 신경 연산의 잠재력을 보여주었습니다.

잠재력을 보여주었습니다.

## 2.2 로젠블랫의 퍼셉트론 모델(1957)

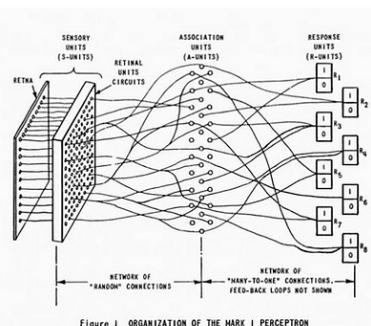
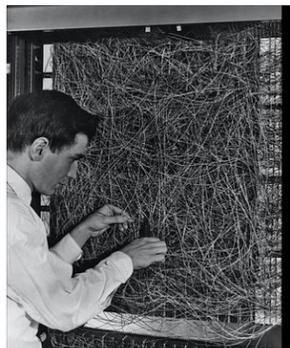
1957년 프랭크 로젠블랫은 패턴을 학습하고 인식할 수 있는 단일 계층 신경망인 퍼셉트론을 소개했습니다. 퍼셉트론 모델은 MP 뉴런보다 더 일반적인 계산 모델로, 실제 값의 입력을 처리하고 가중치를 조정하여 분류 오류를 최소화하도록 설계되었습니다.



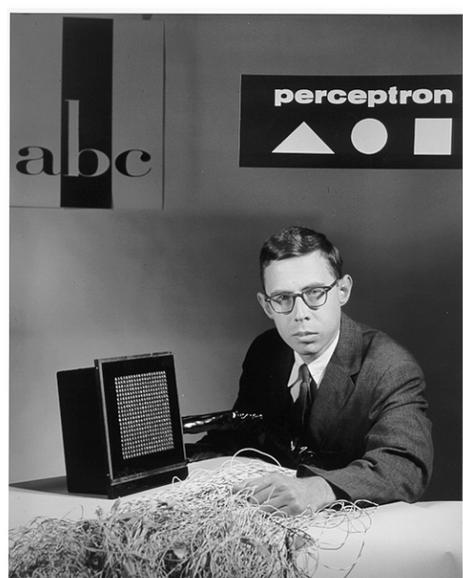
Frank Rosenblatt

또한 로젠블랫은 퍼셉트론을 위한 지도 학습 알고리즘을 개발하여 네트워크가 학습 데이터에서 직접 학습할 수 있도록 했습니다.

$$\mathcal{L}(\mathbf{w}) = - \sum_{i \in \mathcal{M}} \mathbf{w}^T \mathbf{x}_i y_i$$



The Mark I Perceptron was a machine that implemented the perceptron algorithm for image recognition.



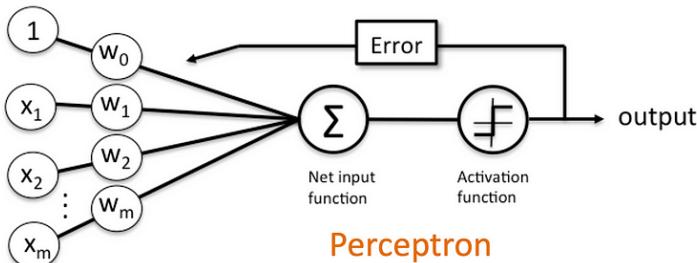
개인을 인식하고 언어 간 음성을 번역할 수 있는 잠재력을 포함한 퍼셉트론의 기능에 대한 로젠블랫의 야심찬 주장은 그 기간 동안 AI에 대한 대중의 관심을 크게 불러일으켰습니다. 퍼셉트론 모델과 관련 학습 알고리즘은 신경망의 진화에 중요한 이정표가 되었습니다.

그러나 곧 퍼셉트론의 학습 규칙이 비선형적으로 분리 가능한 훈련 데이터를 제시할 때 수렴할 수 없다는 중대한 한계가 드러났습니다.

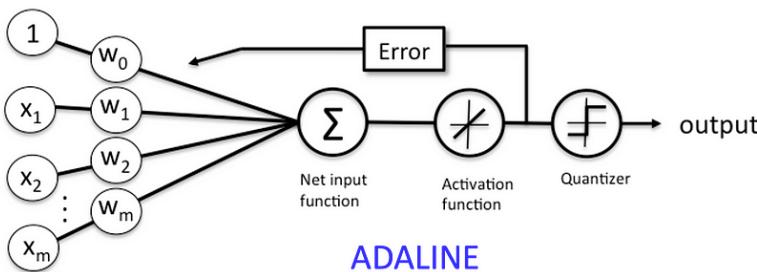
## 2.3 아달린 (1959)

1959년, 위드로와 호프는 퍼셉트론 학습 규칙을 개선한 **ADALINE** (적응형 선형 뉴런, 일명 델타 학습 규칙)을 소개했습니다.

ADALINE은 이진 출력 및 노이즈 감도와 같은 한계를 해결하고 비선형적으로 분리 가능한 데이터를 학습하고 수렴할 수 있어 신경망 개발의 획기적인 발전을 이루었습니다.



Bernard Widrow

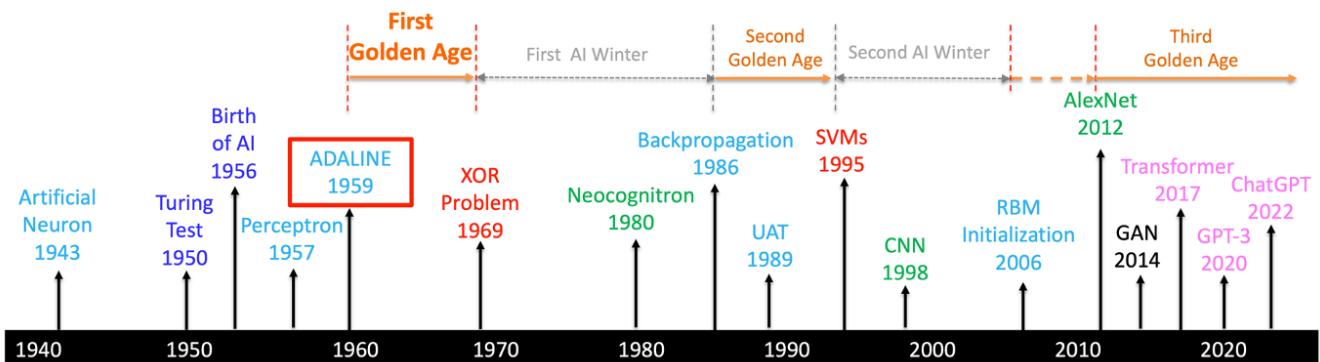


Marcian Hoff

아달린의 주요 기능은 다음과 같습니다:

- 선형 활성화 기능: 퍼셉트론의 스텝 기능과 달리, ADALINE은 선형 활성화 기능을 사용하므로 회귀 작업 및 연속 출력에 적합합니다.
- 최소 평균 제곱(LMS) 알고리즘: ADALINE은 평균 제곱 오차를 최소화하는 LMS 알고리즘을 사용합니다. 평균 제곱 오차를 최소화하여 보다 효율적이고 안정적인 효율적이고 안정적인 학습 프로세스를 제공합니다.
- 적응형 가중치: LMS 알고리즘은 학습 결과의 오차에 따라 출력의 오차에 따라 적응적으로 가중치를 조정하여 ADALINE이 이 노이즈가 있는 상황에서도 효과적으로 학습하고 수렴할 수 있도록 합니다.

아달린의 도입은 로젠블랫의 퍼셉트론 학습의 한계를 극복하고 신경망의 첫 번째 황금기가 시작되는 계기가 되었습니다. 이 획기적인 기술은 효율적인 학습, 지속적인 출력, 노이즈 데이터에 대한 적응을 가능하게 하여 이 분야에서 혁신과 빠른 발전의 물결을 일으켰습니다.



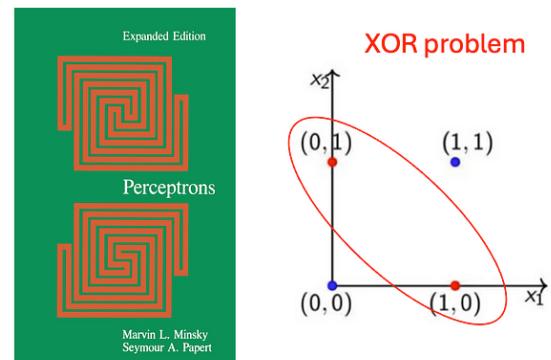
그러나 퍼셉트론과 마찬가지로 아달린은 여전히 선형적으로 분리 가능한 문제에만 국한되어 더 복잡하고 비선형적인 작업을 해결할 수 없었습니다. 이러한 한계는 나중에 XOR 문제로 부각되어 더욱 발전된 신경망 아키텍처의 개발로 이어졌습니다.

## 2.4 XOR 문제(1969)

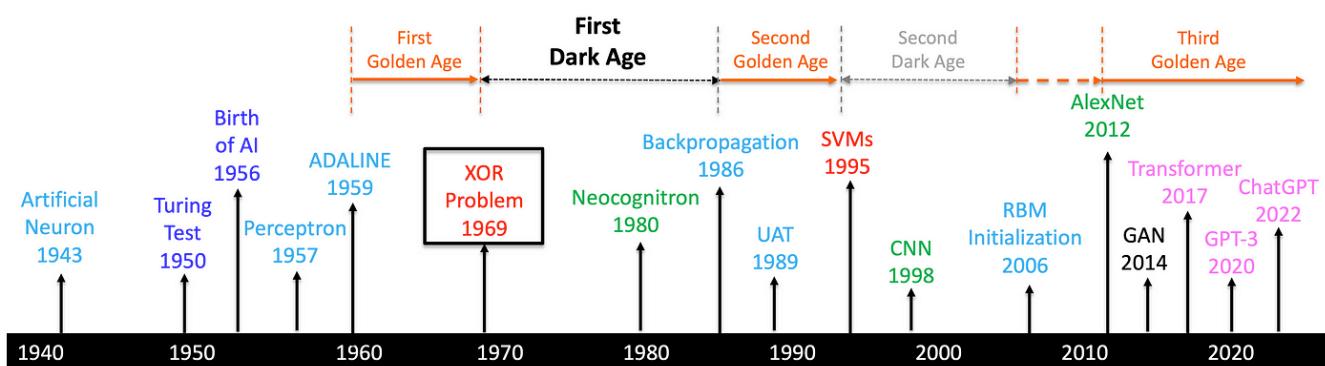
1969년, 마빈 민스키와 시모어 파퍼트는 저서 "퍼셉트론"에서 단층 퍼셉트론의 치명적인 한계를 강조했습니다. 이들은 퍼셉트론이 선형적인 결정 경계로 인해 단순한 이진 분류 작업인 배타적 OR(XOR) 문제를 해결할 수 없음을 보여주었습니다. XOR 문제는 선형적으로 분리할 수 없으므로 단일 선형 경계로는 모든 입력 패턴을 정확하게 분류할 수 없습니다.



**Minsky and Papert:** Perceptrons: An introduction to computational geometry. MIT Press, 1969.



이 발견은 비선형 의사 결정 경계를 학습할 수 있는 보다 복잡한 신경망 아키텍처의 필요성을 강조했습니다. 퍼셉트론의 한계가 드러나면서 신경망에 대한 신뢰를 잃고 기호적 AI 방식으로 전환하게 되었고, 1970년대 초부터 1980년대 중반까지 '신경망의 첫 번째 암흑기'가 시작되었습니다.

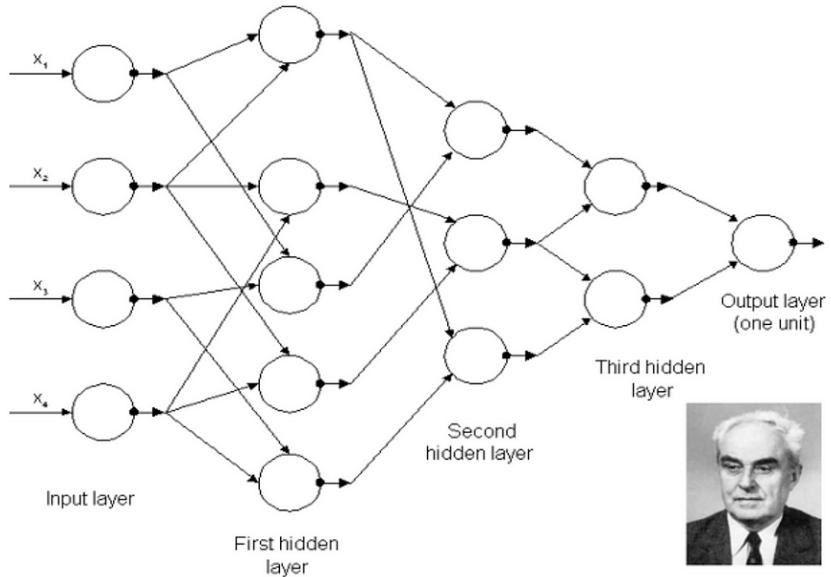


신경망의 첫 번째 암흑기.

그러나 XOR 문제를 해결하면서 얻은 인사이트를 통해 연구자들은 비선형 관계를 포착할 수 있는 더 복잡한 모델의 필요성을 인식하게 되었습니다. 이러한 깨달음은 궁극적으로 다층 퍼셉트론과 기타 고급 신경망 모델의 개발로 이어졌고, 이후 수십 년 동안 신경망과 딥 러닝이 부활하는 발판을 마련했습니다.

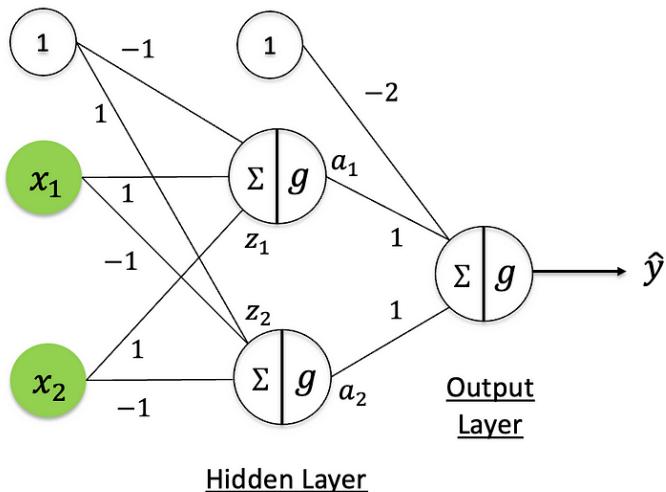
## 3. 다층 퍼셉트론(1960년대)

다층 퍼셉트론(MLP)은 1960년대에 단일 레이어 퍼셉트론을 개선하기 위해 도입되었습니다. 여러 층의 상호 연결된 뉴런으로 구성되어 있어 단일 레이어 모델의 한계를 극복할 수 있습니다. 소련의 과학자 **A. G. 이바크넨코**와 **V. 라파**는 퍼셉트론의 기초 연구를 바탕으로 MLP 개발에 큰 공헌을 했습니다.



### 3.1 Hidden Layer 은닉층

은닉층을 추가하면 MLP가 데이터의 복잡한 비선형 관계를 포착하고 표현할 수 있습니다. 이러한 은닉층은 네트워크의 학습 능력을 크게 향상시켜 XOR 문제와 같이 선형적으로 분리할 수 없는 문제를 해결할 수 있게 해줍니다.



XOR			
$(x_1, x_2)$	$(z_1, z_2)$	$(a_1, a_2)$	$\hat{y}$
(0, 0)	(-1, 1)	(0, 1)	0
(0, 1)	(0, 0)	(1, 1)	1
(1, 0)	(0, 0)	(1, 1)	1
(1, 1)	(1, -1)	(1, 0)	0

$a = g(z) = u(z) = \begin{cases} 1, & z \geq 0 \\ 0, & \text{otherwise} \end{cases}$

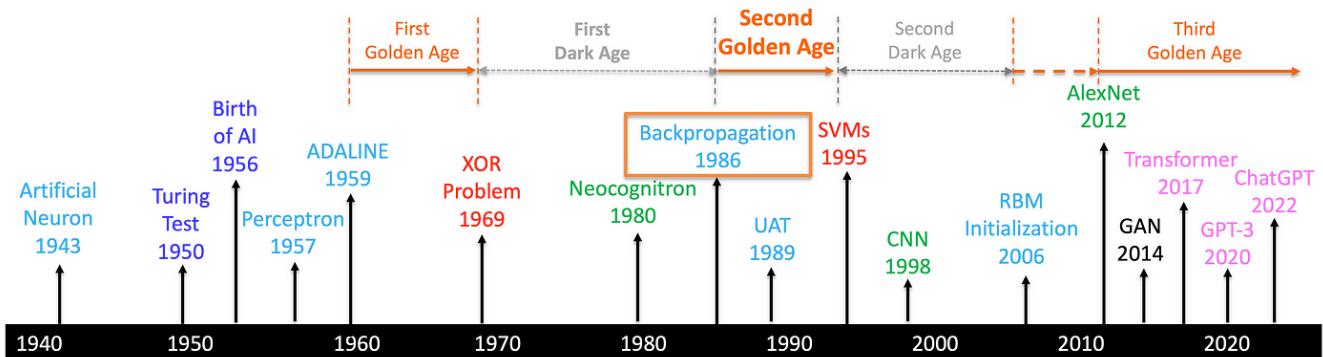
Activation function is the unit step function  $u(z)$

### 3.2 MLP의 역사적 맥락과 과제

MLP는 복잡한 문제를 해결하기 위한 딥 러닝 아키텍처의 잠재력을 보여주며 신경망 연구의 획기적인 발전을 이루었습니다. 하지만 1960년대와 1970년대에는 몇 가지 문제로 인해 MLP의 발전이 저해되었습니다:

- 학습 알고리즘의 부족: 초기 MLP 모델에는 효율적인  
네트워크의 가중치를 효과적으로 조정할 수 있는 훈련 알고리즘이  
네트워크의 가중치를 효과적으로 조정할 수 있는 효율적인 훈련 알고리즘이 부족했습니다. 역전파가 없었기 때문에 여러 계층으로 구성된 딥 네트워크를  
여러 계층으로 구성된 심층 네트워크를 훈련하는 것이 어려웠습니다.
- 계산 능력의 한계: 당시 사용 가능한 컴퓨터 파워  
딥 뉴럴 네트워크 훈련에 필요한 복잡한 계산을 처리하기에는  
복잡한 계산을 처리하기에 충분하지 않았습니다. 이 한계는  
MLP 연구 및 개발의 진전을 더디게 했습니다.

신경망의 첫 번째 암흑기는 1986년 역전파 알고리즘의 재발견과 발표로 끝났고, 신경망의 두 번째 황금기가 시작되었습니다.



신경망의 두 번째 황금기.

## 4. Backpropagation 역전파(1970년대-1980년대)

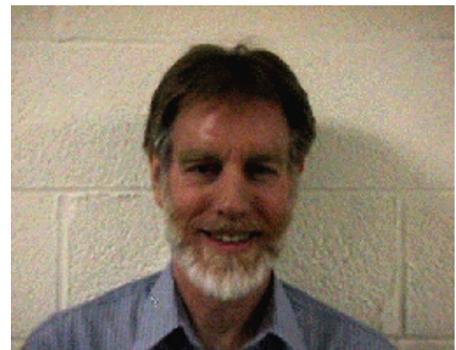
1969년, XOR 문제는 퍼셉트론(단층 신경망)의 한계를 부각시켰습니다. 연구자들은 다층 신경망이 이러한 한계를 극복할 수 있다는 사실을 깨달았지만, 이러한 복잡한 네트워크를 훈련할 수 있는 실용적인 알고리즘이 부족했습니다. 신경망이 이론적으로 모든 함수에 근사치를 구할 수 있는 역전파 알고리즘이 개발되기까지 17년이 걸렸습니다. 흥미롭게도 이 알고리즘은 실제로 발표되기 전에 이미 발명되었다는 사실이 나중에 밝혀졌습니다. 오늘날 역전파는 1960년대와 1970년대에 처음 등장한 이후 상당한 발전과 개선을 거쳐 딥러닝의 기본 구성 요소로 자리 잡았습니다.



**David Rumelhart**



**Geoffrey Hinton**



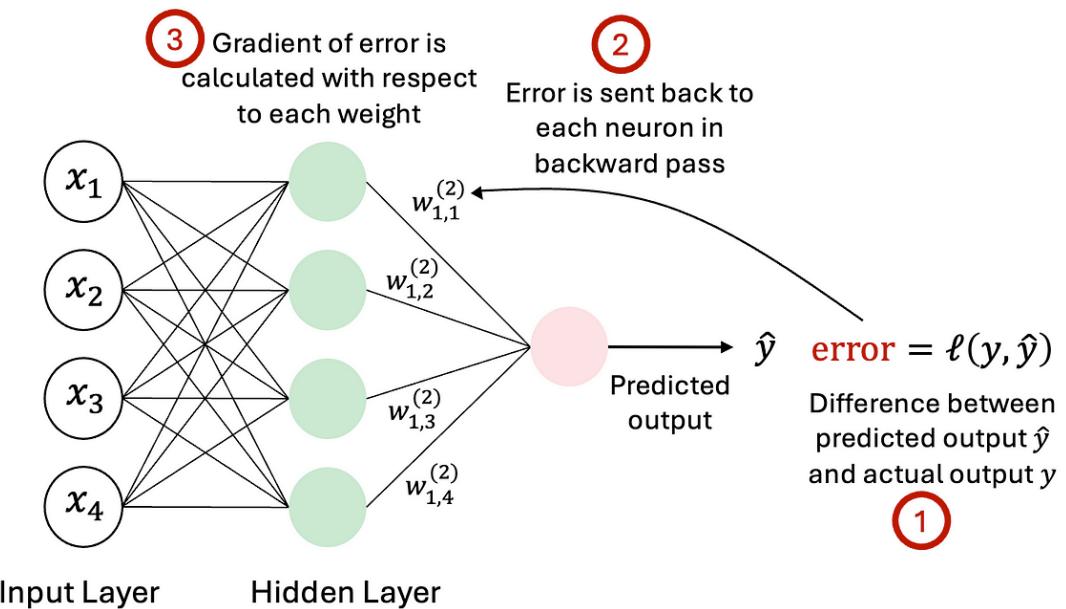
**Ronald Williams**

### 4.1 초기 개발(1970년대)

- 세포 린나인마(**1970**): 역전파 알고리즘의 핵심 구성 요소인 자동 차별화 개념을 도입했습니다.
- 폴 베르보스(**1974**): 미적분학의 연쇄 규칙을 사용하여 네트워크의 가중치에 대한 오차 함수의 기울기를 계산하여 다층 신경망의 학습을 가능하게 하는 방법을 제안했습니다.

### 4.2 개선과 대중화(1980년대)

- 데이비드 럼멜하트, 제프리 힌튼, 로널드 윌리엄스(**1986**): 심층 신경망을 훈련하기 위한 실용적이고 효율적인 방법으로 역전파를 제시하고 다양한 문제에 적용하는 방법을 시연했습니다.



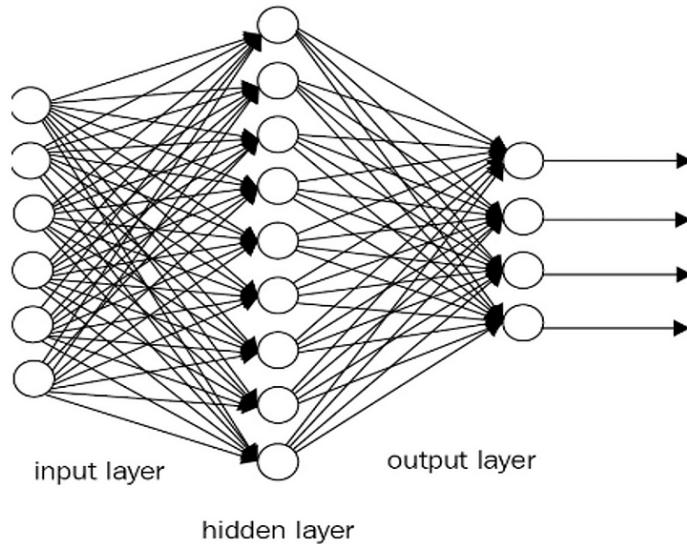
### 4.3 역전파의 주요 기능:

- Gradient Descent** 그라데이션 하강: 역전파는 오차 함수를 최소화하기 위해 경사 하강과 함께 사용됩니다. 이 알고리즘은 네트워크의 각 가중치에 대한 오차의 기울기를 계산하여 가중치를 반복적으로 업데이트하여 오차를 줄일 수 있도록 합니다.
- Chain Rule** 체인 규칙: 역전파 알고리즘의 핵심은 미적분학의 연쇄 규칙을 적용하는 것입니다. 이 규칙을 사용하면 오차의 기울기를 일련의 부분 미분으로 분해할 수 있으며, 네트워크를 통한 역전파를 통해 효율적으로 계산할 수 있습니다.
- Layer Computation** 계층화된 계산: 역전파는 출력 레이어에서 시작하여 입력 레이어로 역방향으로 작동하는 레이어별 방식으로 작동합니다. 이 계층화된 계산은 그래디언트가 네트워크를 통해 올바르게 전파되도록 하여 심층 아키텍처를 학습할 수 있게 해줍니다.

### 4.5 ## Universal Approximation Theorem(1989)

1989년 조지 사이벤코가 제안한 범용 근사 정리는 다층 신경망의 기능에 대한 수학적 토대를 제공했습니다. 이 정리에 따르면 단일 숨겨진 계층을 가진 피드포워드 신경망은 충분한 뉴런이 주어지고 비선형 활성화 함수를 사용하면 모든 연속 함수를 임의의 정확도로 근사화할 수 있습니다. 이 정리는 신경망의 강력한 성능과 유연성을 강조하여 다양한 애플리케이션에 적합합니다.

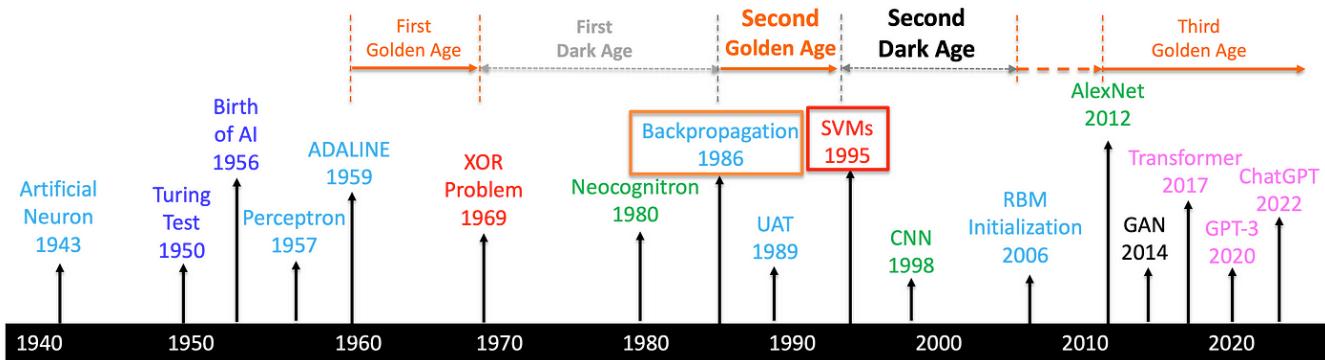
#### One Hidden Layer is Enough



단일 숨겨진 계층을 가진 다층 신경망은 모든 연속 함수를 원하는 정밀도로 근사화할 수 있어 다양한 영역에서 복잡한 문제 해결을 가능하게 합니다.

### 4.6 제2의 황금기(1980년대 후반~1990년대 초반)

역전파와 UAT(범용 근사 정리)의 개발은 신경망의 두 번째 황금기가 시작되는 계기가 되었습니다. 역전파는 다층 신경망을 훈련하는 효율적인 방법을 제공하여 연구자들이 더 깊고 복잡한 모델을 훈련할 수 있게 했습니다. UAT는 다층 신경망 사용에 대한 이론적 정당성을 제공하고 복잡한 문제를 해결할 수 있는 능력에 대한 신뢰를 강화했습니다. 1980년대 후반과 1990년대 초반에 걸친 이 시기에는 이 분야에 대한 관심이 다시 높아지고 상당한 발전이 이루어졌습니다.



## 4.7 두 번째 암흑기(1990년대 초반~2000년대 초반)

그러나 신경망 분야는 여러 가지 요인으로 인해 1990년대 초부터 2000년대 초까지 '제2의 암흑기'를 경험했습니다:

- 분류 및 회귀 작업에 수학적으로 우아한 접근 방식을 제공하는 서포트 벡터 머신(**SVM**)이 등장했습니다.
- 심층 신경망을 훈련하는 데 여전히 시간이 많이 걸리고 하드웨어 집약적이었던 계산적 한계.
- 과적합 및 일반화 문제: 초기 신경망은 훈련 데이터에서는 잘 작동하지만 보이지 않는 데이터에서는 성능이 저하되어 실제 적용에 대한 신뢰성이 떨어졌습니다.

이러한 어려움으로 인해 많은 연구자들이 신경망에 대한 관심을 다른 분야로 옮기게 되었고, 이로 인해 신경망 분야는 정체기에 접어들게 되었습니다.

## 4.8 딥러닝의 부활(2000년대 후반~현재):

신경망 분야는 2000년대 후반과 2010년대 초반에 신경망의 발전으로 인해 다시 부흥을 경험했습니다:

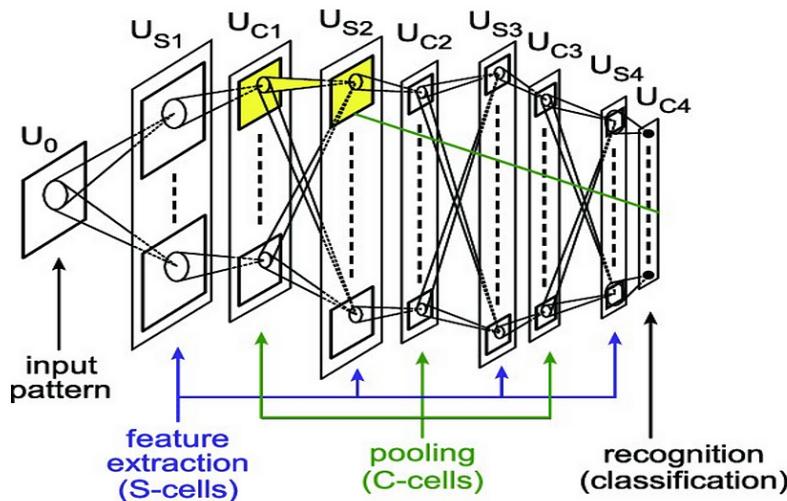
- 딥 러닝 아키텍처 (CNN, RNN, 트랜스포머, 확산 모델)
- 하드웨어 (GPU, TPU, LPU)
- 대규모 데이터 세트 (이미지넷, 코코, 오픈웹)
- 텍스트, 위키텍스트 등)
- 훈련 알고리즘 (SGD, Adam, 드롭아웃)

이러한 발전은 컴퓨터 비전, 자연어 처리, 음성 인식, 강화 학습 분야에서 중요한 돌파구를 마련했습니다. 범용 근사 정리는 실용적인 발전과 결합하여 딥러닝 기술이 널리 채택되고 성공할 수 있는 길을 열었습니다.

## 5. 컨볼루션 신경망(1980년대~2010년대)

컨볼루션 신경망(CNN)은 특히 컴퓨터 비전과 이미지 처리 분야에서 딥 러닝의 지형을 극적으로 변화시켰습니다. 1980년대부터 2010년대까지의 발전은 아키텍처, 학습 기술 및 애플리케이션의 상당한 발전을 반영합니다.

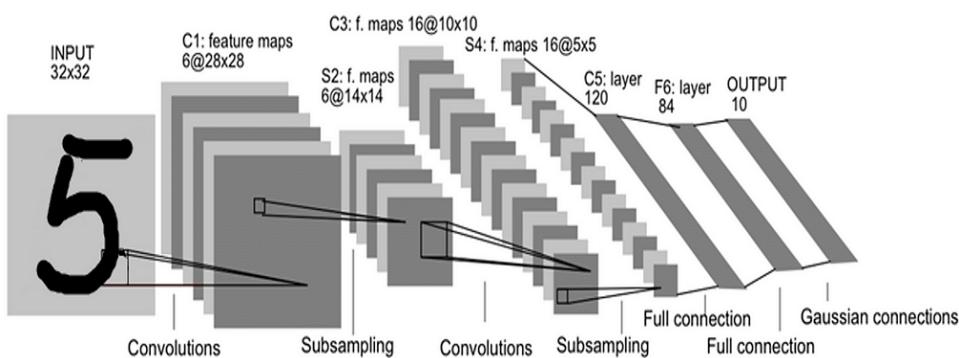
# Neocognitron



Kenji Fukushima

## 5.1 초기 개발(1989-1998)

CNN의 개념은 1980년대에 인간의 시각 피질 구조를 모방한 계층적 신경망인 네오코그니트론을 제안한 후쿠시마 켄지에 의해 처음 소개되었습니다. 이 선구적인 연구는 CNN 개발의 토대를 마련했습니다. 1980년대 말과 1990년대 초에 얀 르쿤과 그의 팀은 손으로 쓴 숫자 인식을 위해 특별히 설계된 LeNet-5 아키텍처를 도입하여 CNN을 더욱 발전시켰습니다.



Yann LeCun

## 5.2 CNN의 주요 구성 요소

CNN은 세 가지 주요 구성 요소로 구성됩니다:

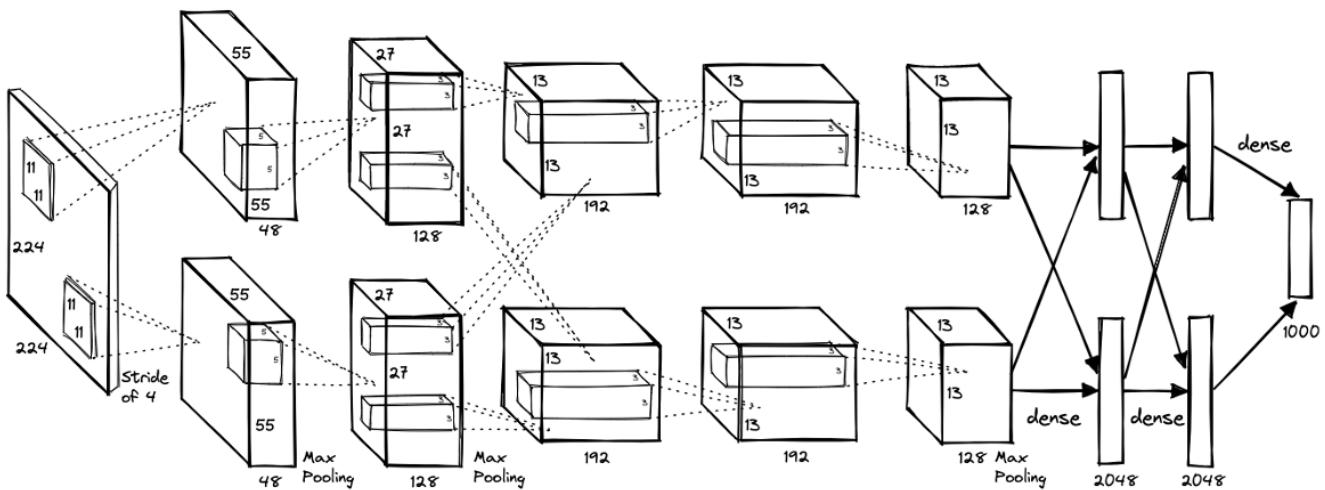
1. 컨볼루션 레이어: 이 레이어는 학습 가능한 필터 세트를 적용하여 입력 이미지에서 특징의 공간 계층 구조를 자동으로 학습합니다.
2. 풀링 레이어: 풀링 레이어는 입력의 공간적 차원을 줄여 변화에 대한 견고성을 높이고 계산 부하를 줄입니다.
3. 완전히 연결된 레이어: 컨볼루션 및 풀링 레이어에 이어 완전히 연결된 레이어는 분류 작업에 사용되며, 이전 레이어에서 학습한 특징을 통합합니다.

## 5.3 CNN의 주요 기능

- **로컬 수신 필드:** CNN은 로컬 수신 필드를 사용하여 입력 데이터의 로컬 패턴을 캡처하므로 이미지 및 시각 작업에 매우 효과적입니다.
- **공유 가중치:** 컨볼루션 레이어에서 공유 가중치를 사용하면 네트워크의 매개변수 수가 줄어들어 더 효율적이고 쉽게 학습할 수 있습니다.
- **번역 불변성:** 풀링 레이어는 번역 불변성을 도입하여 네트워크가 입력 이미지의 위치에 관계없이 패턴을 인식할 수 있도록 합니다.

## 5.4 CNN의 부상: 알렉스넷의 영향력 (2012)

2012년에는 이미지넷 대규모 시각 인식 챌린지(ILSVRC)에서 AlexNet이 큰 점수 차로 승리하며 이미지 분류 분야에서 획기적인 발전을 이루면서 CNN 개발의 중요한 이정표가 세워졌습니다.



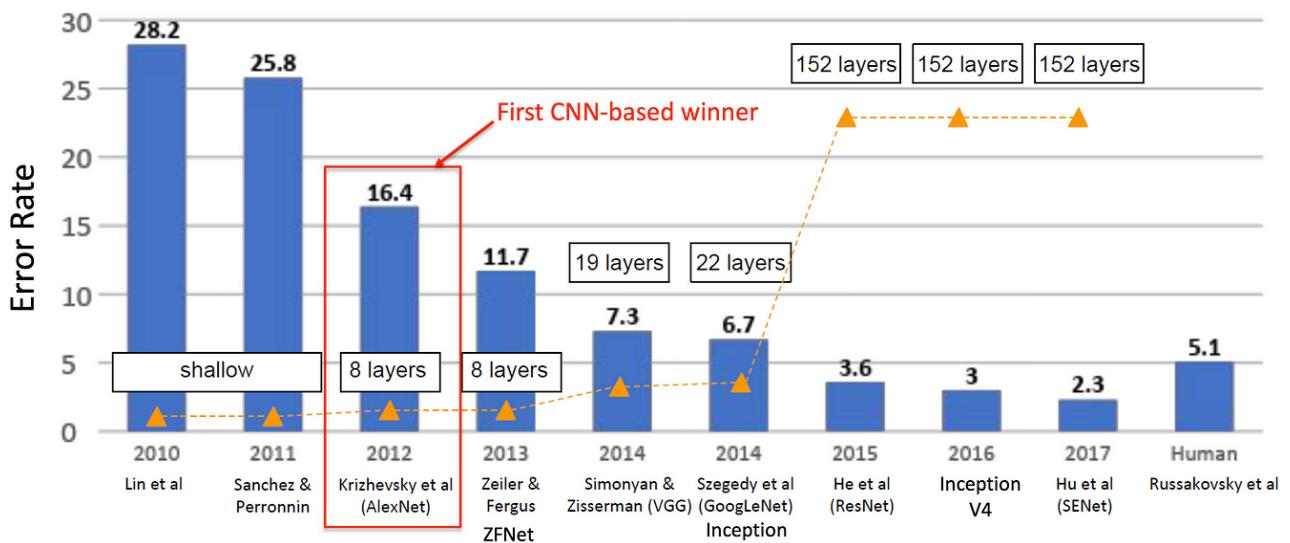
알렉스넷의 아키텍처 (2012)

ILSVRC는 매년 1000개 클래스로 분류된 1000만 개 이상의 주석이 달린 이미지 데이터 세트에 대한 알고리즘을 평가하는 이미지 인식 벤치마크입니다. AlexNet의 혁신이 포함되었습니다:

1. ReLU 활성화 기능: 기존 활성화 기능의 문제를 극복하기 위해 도입된 ReLU는 더 빠른 트레이닝을 가능하게 하고 성능을 개선했습니다.
2. 드롭아웃 정규화: 이 기술은 훈련 중에 무작위로 유닛을 탈락시켜 과적합을 줄였습니다.
3. 데이터 증강: 훈련 데이터 세트의 개선으로 훈련 데이터의 다양성을 인위적으로 증가시켜 일반화를 개선했습니다.

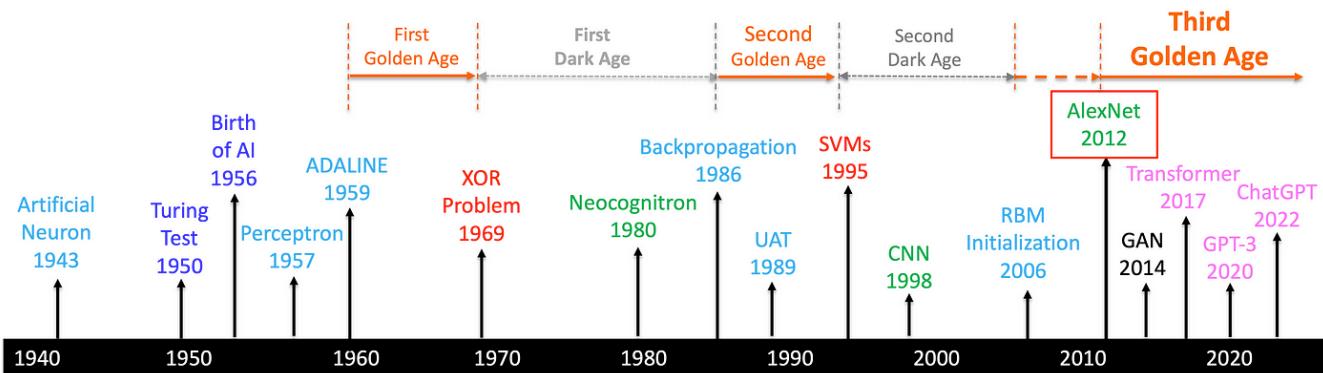
AlexNet의 성공은 CNN 개발의 전환점이 되어 이미지 분류와 객체 감지를 더욱 발전시킬 수 있는 기반을 마련했습니다.

## Winners of ImageNet Classification Challenge (ILSVRC)



## 신경망의 세 번째 황금기를 여는 AlexNet

현재의 황금기(2010년대~현재)는 딥러닝, 빅 데이터, 강력한 컴퓨팅 플랫폼의 융합으로 특징지어집니다. 이 시대에는 이미지 인식, 자연어 처리, 로봇 공학 분야에서 괄목할 만한 혁신이 이루어졌습니다. 현재 진행 중인 연구는 AI 기능의 한계를 계속 넓혀가고 있습니다.



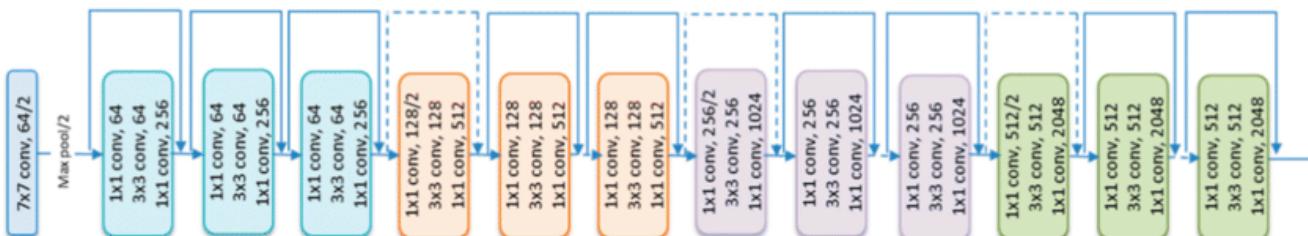
AlexNet, 신경망의 세 번째 황금기를 열다.

## 5.5 후속 아키텍처

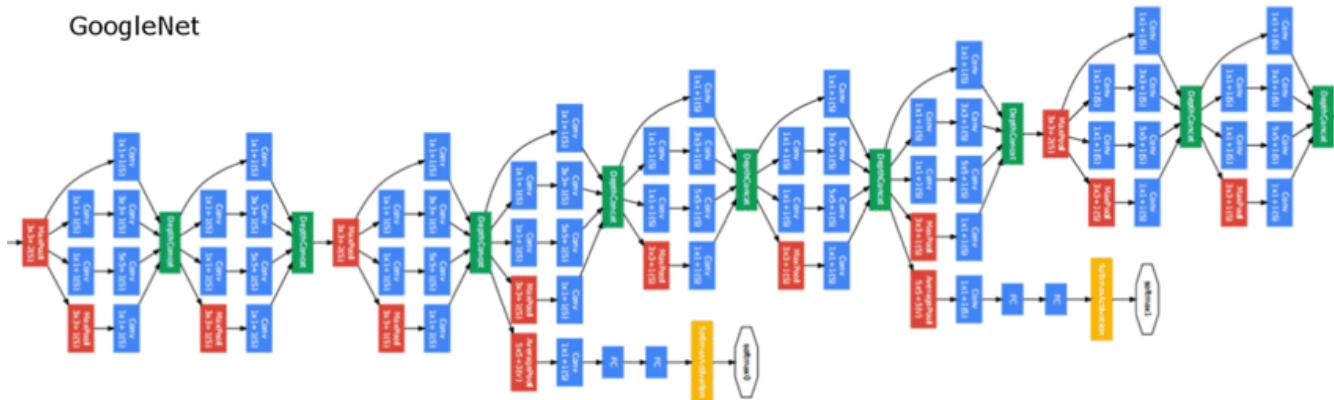
AlexNet 이후 몇 가지 영향력 있는 아키텍처가 등장했습니다:

- **VGGNet (2014):** 옥스퍼드의 시각 기하학 그룹에서 개발한 VGGNet은 더 작은 컨볼루션 필터( $3 \times 3$ )로 더 깊은 아키텍처를 강조하여 놀라운 정확도를 달성했습니다.
- 구글렛/인셉션(2014): 네트워크가 다중 규모의 특징을 효율적으로 캡처할 수 있는 인셉션 모듈을 도입했습니다.
- **ResNet(2015):** 잔여 네트워크는 연결 건너뛰기를 도입하여 매우 심층적인 네트워크를 훈련하는 동시에 소실 경사 문제를 완화할 수 있게 되었습니다.

ResNet



GoogleNet



ResNet 및 GoogLeNet 아키텍처

## 5.6 CNN의 애플리케이션

CNN의 발전은 다양한 분야에 혁신을 가져왔습니다:

- **컴퓨터 비전:** CNN은 현대 컴퓨터 비전의 중추가 되어 이미지 분류, 물체 감지, 의미론적 세분화 분야에서 획기적인 발전을 이루었습니다.
- **의료 영상:** CNN은 질병 진단, 종양 탐지, 이미지 유도 수술과 같은 작업에 활용되어 진단 정확도를 크게 향상시킵니다.
- **자율 주행 차량:** CNN은 자율 주행 자동차의 인식 시스템에 필수적인 요소로, 주변 환경을 해석하고 이에 대응할 수 있게 해줍니다.



CNN의 컴퓨터 비전 애플리케이션.

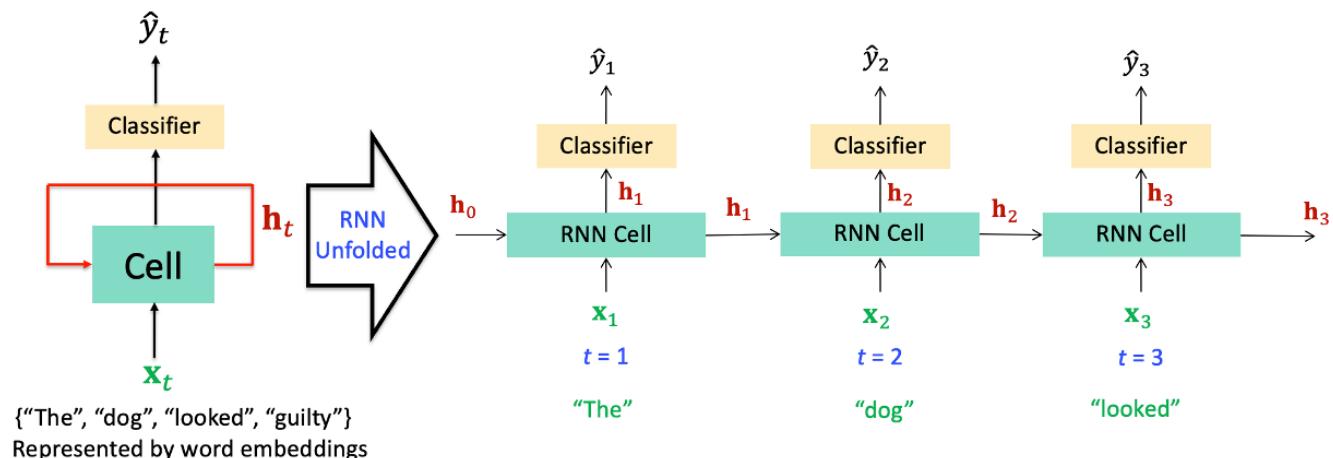
CNN의 시작부터 딥 러닝의 초석이 된 현재까지의 여정은 CNN이 AI에 미친 지대한 영향을 잘 보여줍니다. CNN의 성공은 또한 딥 러닝의 발전을 위한 길을 열었으며, RNN과 트랜스포머와 같은 다른 전문 신경망 아키텍처의 개발에 영감을 주었습니다. CNN의 이론적 토대와 실용적 혁신은 다양한 영역에서 딥러닝 기술이 널리 채택되고 성공하는 데 크게 기여했습니다.

## 6. 순환 신경망 (1986-2017)

순환신경망(RNN)은 순차적이고 시간적인 데이터를 처리하기 위한 강력한 아키텍처로 등장했습니다. 피드포워드 신경망과 달리 RNN은 일련의 입력을 처리하도록 설계되어 언어 모델링, 시계열 예측, 음성 인식과 같은 작업에 특히 효과적입니다.

### 6.1 초기 개발(1980~1990년대)

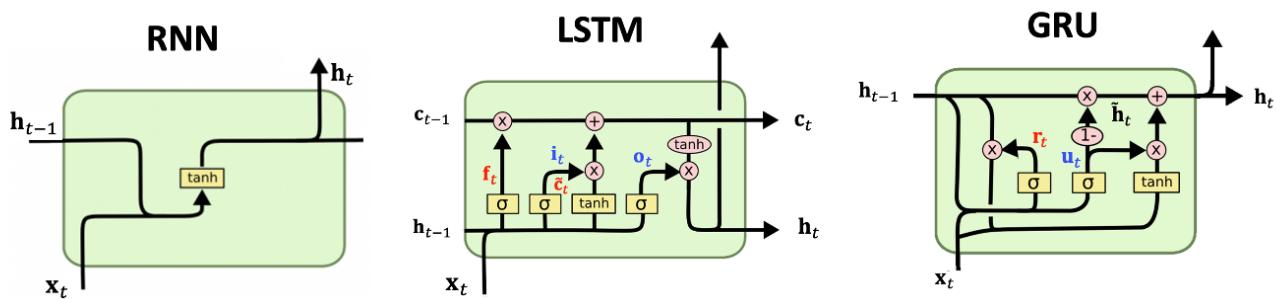
RNN의 개념은 1980년대로 거슬러 올라가며, 존 휙필드, 마이클 조던, 제프리 엘먼과 같은 선구자들이 이 네트워크의 개발에 기여했습니다. 1982년 존 휙필드가 소개한 휙필드 네트워크는 신경망의 반복적 연결을 이해하는 토대를 마련했습니다. 1980년대와 1990년대에 각각 제안된 조던 네트워크와 엘먼 네트워크는 순차적 데이터에서 시간적 의존성을 포착하려는 초기 시도였습니다.



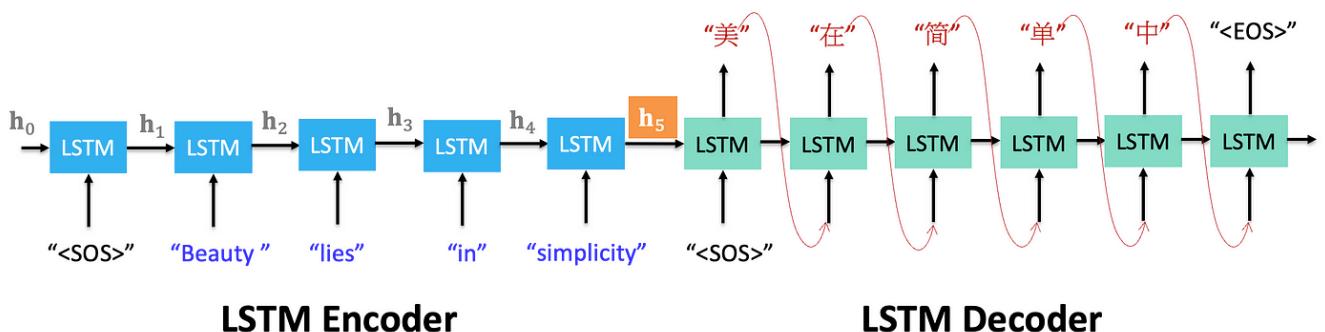
### 6.2 LSTM, GRU 및 Seq2Seq 모델(1997~2014년)

- 장단기 기억(LSTM) 네트워크(1997): 셉 호크라이터(Sepp Hochreiter)와 위르겐 슈미트후버(Jürgen Schmidhuber)는 기존 RNN의 사라지는 기울기 문제를 해결한 LSTM(Long ShortTerm Memory) 네트워크를 소개했습니다. LSTM은 게이팅 메커니즘을 사용해 정보의 흐름을 제어함으로써 순차적인 데이터에서 장기적인 종속성을 포착할 수 있습니다.

- **게이트 순환 유닛(GRU)(2014)**: 조경현 등은 정보의 흐름을 제어하기 위해 게이팅 메커니즘을 사용하는 LSTM의 단순화된 버전인 GRU(Gated Recurrent Unit)를 제안했습니다. GRU는 LSTM보다 파라미터 수가 적고 훈련 속도가 더 빠른 경우가 많습니다.



- **시퀀스 대 시퀀스 모델(Seq2Seq) (2014)**: 일리야 수츠케버와 그의 팀은 인코더-디코더 아키텍처를 사용하여 입력 시퀀스를 출력 시퀀스에 매핑하는 Seq2Seq 모델을 소개했습니다. 이 모델은 기계 번역, 음성 인식, 텍스트 요약과 같은 작업에 널리 사용되었습니다.



## 6.3 RNN의 주요 특징 반복 연결:

RNN은 반복 연결을 사용하여 이전 시간 단계의 정보를 캡처하는 숨겨진 상태를 유지합니다. 이를 통해 네트워크는 순차적 데이터의 시간적 종속성을 모델링할 수 있습니다.

- **시간을 통한 역전파(BPTT):** RNN은 시간에 따라 반복 네트워크를 전개하고 전개된 네트워크에 표준 역전파 알고리즘을 적용하는 시간을 통한 역전파(BPTT)라는 변형된 역전파를 사용하여 학습됩니다.
- **게이팅 메커니즘:** LSTM 및 GRU와 같은 고급 RNN 아키텍처는 게이팅 메커니즘을 사용해 정보의 흐름을 제어함으로써 소실 경사 문제를 완화하고 네트워크가 장기적인 종속성을 포착할 수 있도록 지원합니다.

A screenshot of the Google Translate interface showing an English sentence being translated into Chinese:

English input: Deep learning is extremely powerful.

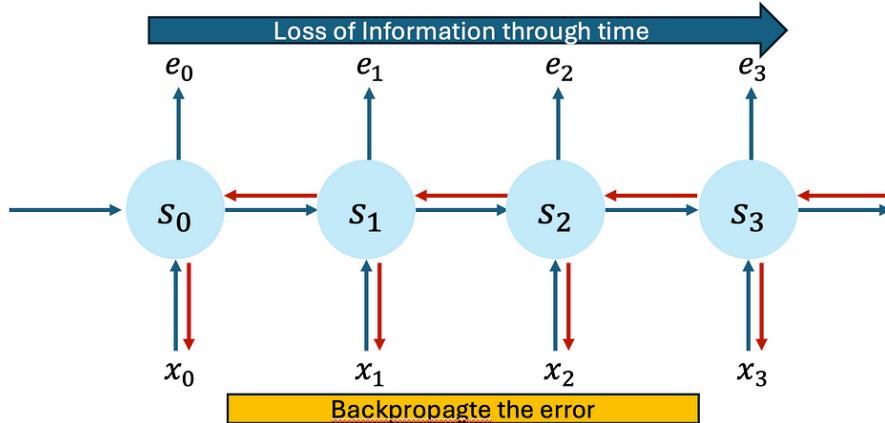
Chinese output: 深度学习非常强大.

Below the output, the pinyin transcription is shown: Shēndù xuéxí fēicháng qíángdà.

## 6.4 RNN 애플리케이션

RNN은 다음과 같은 다양한 분야에 큰 영향을 미쳤습니다:

1. 자연어 처리: RNN은 자연어 처리 분야에 혁명을 일으켜 언어 모델링, 기계 번역, 감정 분석, 텍스트 생성 등의 작업에서 상당한 발전을 이루었습니다.
2. 음성 인식: RNN은 음성 인식 시스템에서 널리 사용되며, 음성 언어의 시간적 종속성을 모델링하여 음성 신호를 텍스트로 변환합니다.
3. 시계열 예측: RNN은 순차적 데이터의 시간적 종속성을 모델링하여 미래 값을 예측하는 시계열 예측에 효과적입니다.



## 6.5 RNN의 과제

이러한 성공에도 불구하고 RNN은 몇 가지 도전 과제에 직면해 있습니다:

- 사라지는 그라데이션과 폭발하는 그라데이션: 기존 RNN은 이러한 문제로 어려움을 겪지만, LSTM과 GRU는 몇 가지 해결책을 제공합니다.
- 계산 복잡성: RNN 훈련은 특히 대규모 데이터 세트의 경우 리소스 집약적일 수 있습니다.
- 병렬화: RNN의 순차적 특성으로 인해 병렬 학습 및 추론 프로세스가 복잡해집니다.

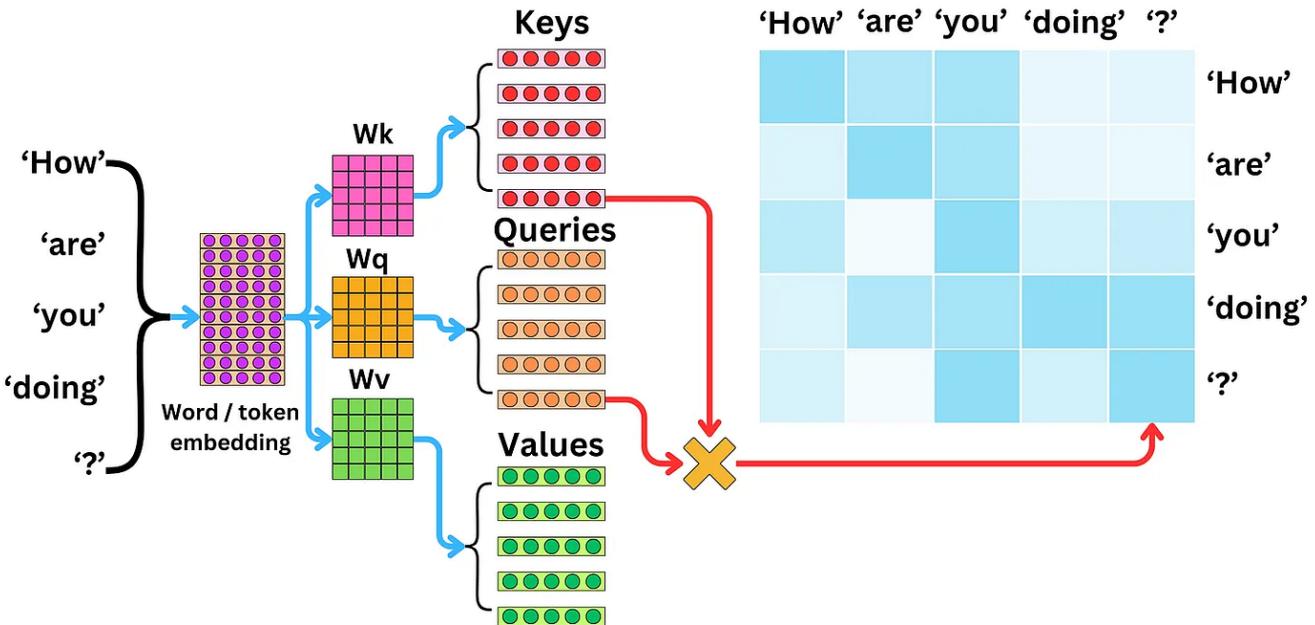
RNN의 성공은 딥 러닝을 더욱 발전시킬 수 있는 길을 열었으며, 다양한 순차적 데이터 작업에서 최첨단 성능을 달성한 트랜스포머와 같은 다른 전문 신경망 아키텍처의 개발에도 영감을 주었습니다. RNN의 이론적 토대와 실용적 혁신은 다양한 영역에서 딥러닝 기술이 널리 채택되고 성공하는 데 크게 기여했습니다.

## 7. 트랜스포머 (2017-현재)

트랜스포머는 순차적 데이터를 처리하는 탁월한 능력으로 딥 러닝의 환경을 변화시켜 자연어 처리(NLP)에서 컴퓨터 비전에 이르기까지 다양한 분야에서 중추적인 역할을 담당하고 있습니다.

### 7.1 트랜스포머 소개 (2017)

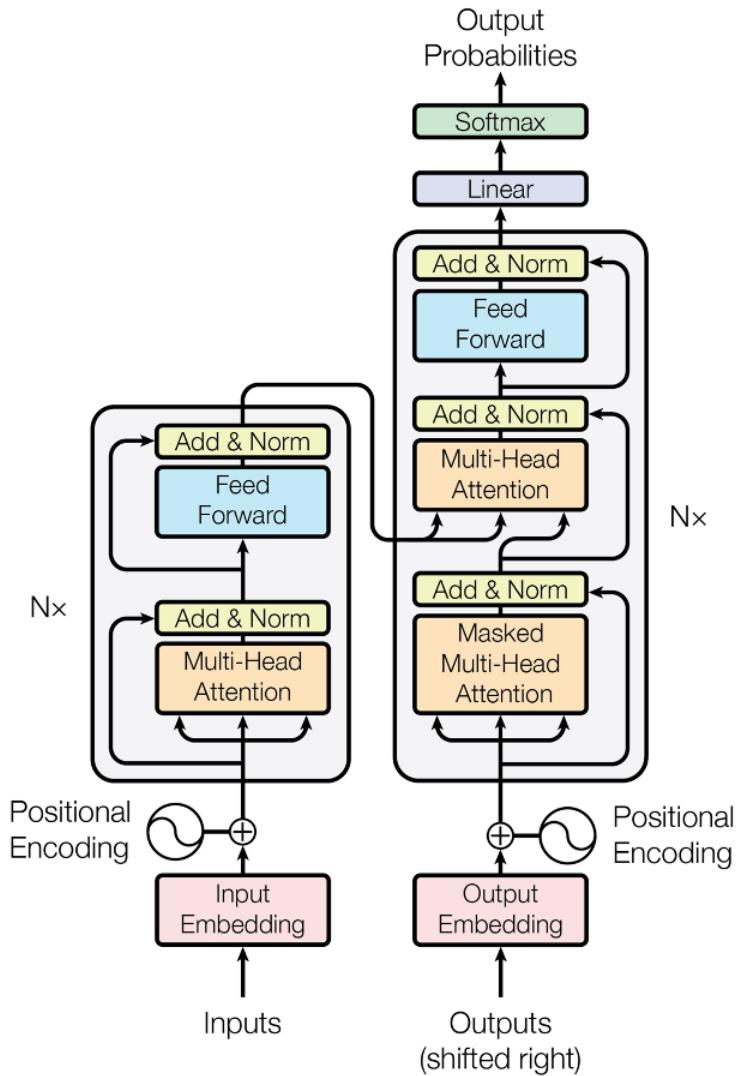
트랜스포머 모델은 Vaswani 외 (2017)가 "attention만 있으면됩니다."라는 논문에서 소개했습니다. 이 모델은 기존의 순차적 RNN 처리 방식을 버리고 **Self- attention** 메커니즘을 채택하여 병렬 처리를 가능하게 하고 장거리 종속성을 더 잘 처리할 수 있게 했습니다.



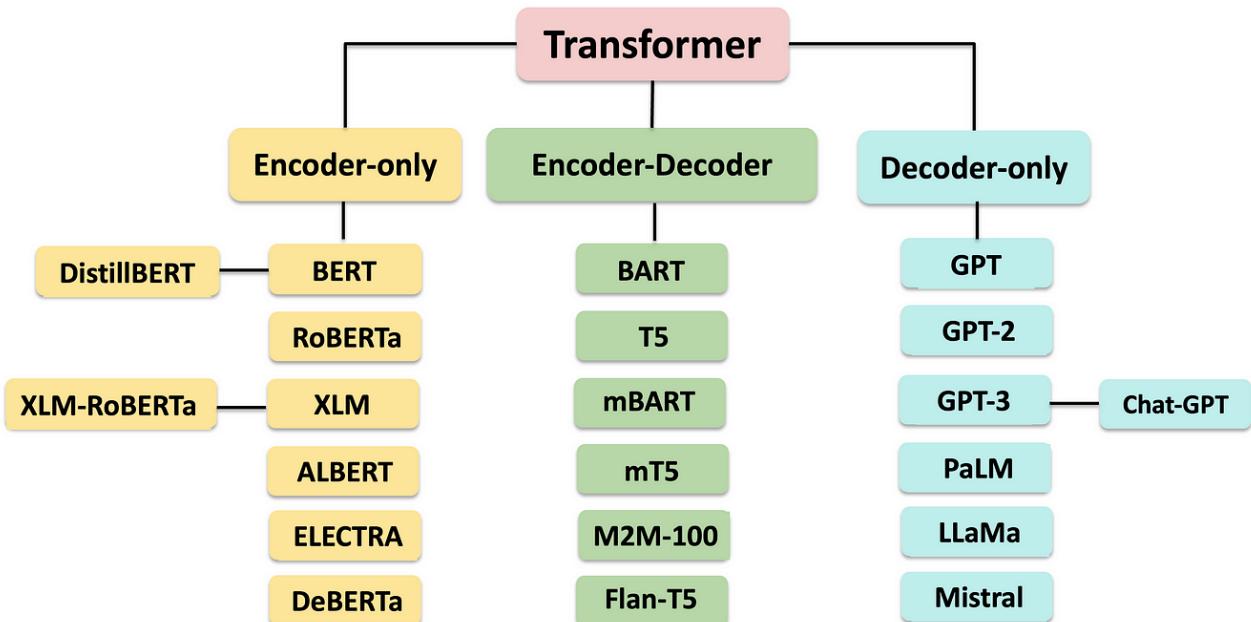
셀프 어텐션 메커니즘([출처](#))

## 7.2 트랜스포머의 주요 기능

- **셀프 어텐션 메커니즘:** 시퀀스의 각 위치가 모든 위치에 attention를 기울여 RNN이나 LSTM보다 더 유연하게 컨텍스트를 캡처할 수 있습니다.
- **병렬화:** 모든 입력 데이터를 동시에 처리하여 훈련 속도를 향상시키며, RNN의 순차적 특성과는 극명한 대조를 이룹니다.
- **인코더-디코더 구조:** 인코더와 디코더 스택 모두 시퀀스 순서를 유지하기 위해 위치 인코딩과 함께 자기 attention 및 피드 포워드 신경망 레이어를 활용합니다.



인코더-디코더 구조와 멀티 헤드 어텐션이 적용된 독창적인 트랜스포머 아키텍처.

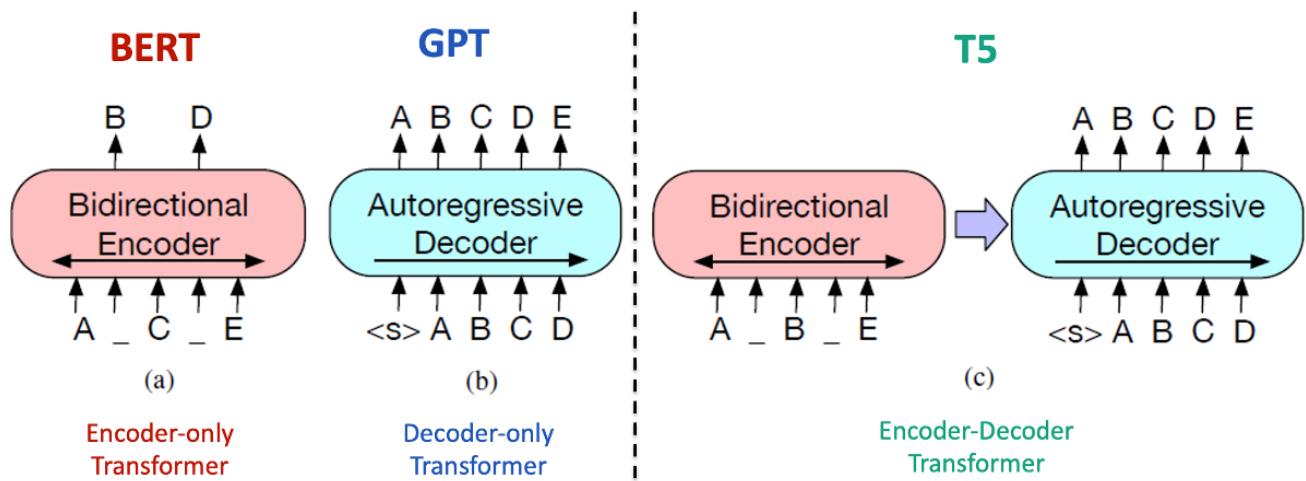


트랜스포머 기반 언어 모델

### 7.3 트랜스포머 기반 언어 모델(2017 - 현재)

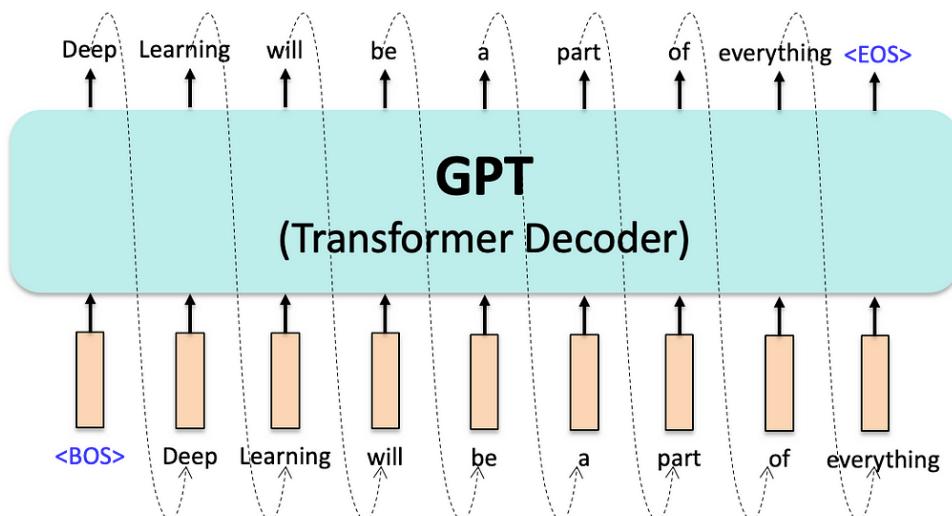
- **BERT (2018):** 인코더 전용 트랜스포머인 Transformers의 양방향 인코더 표현은 마스크 언어 모델링과 다음 문장 예측에 대한 사전 학습을 통해 NLP에 혁신을 일으켰습니다.

- **T5 (2019):** 인코더-디코더 트랜스포머인 텍스트-텍스트 전송 트랜스포머는 NLP 작업을 텍스트-텍스트 형식으로 재구성하여 모델 아키텍처와 학습을 간소화했습니다.



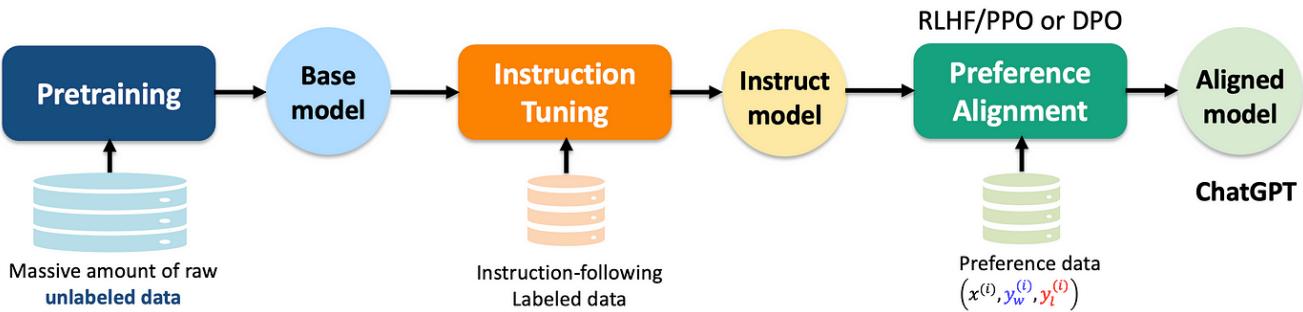
## OpenAI의 GPT 시리즈:

- **GPT (2018):** OpenAI가 도입한 자동 회귀 디코더 전용 트랜스포머인 생성형 사전 훈련 트랜스포머는 텍스트 시퀀스에서 다음 단어를 예측하는 데 중점을 두어 인상적인 언어 이해 및 생성 기능을 보여주었습니다.
- **GPT-2(2019):** 이전 버전보다 훨씬 더 큰 규모로, 제로샷 작업 수행과 같은 새로운 기능을 선보이며 일관된, 때로는 오해의 소지가 있는 텍스트를 생성하는 능력으로 인해 AI의 오용 가능성에 대한 논의를 불러일으켰습니다.
- **GPT-3(2020):** 1,750억 개의 파라미터를 갖춘 GPT-3는 언어 모델에서 가능한 범위를 더욱 확장하여 최소한의 미세 조정으로 작업을 수행하는 데 탁월하며, 이를 '소수 샷 학습(few-shot learning)'이라고 합니다. 디코더 전용 트랜스포머인 GPT-3의 자동 회귀 아키텍처를 통해 시퀀스의 이전 단어에 따라 한 번에 한 단어씩 텍스트를 생성할 수 있습니다.



GPT의 자동 회귀 언어 모델 아키텍처는 입력된 이전 토큰을 기반으로 시퀀스의 다음 토큰을 예측하도록 설계되었습니다.

- **ChatGPT (2022):** 대화 참여에 최적화된 GPT-3.5 시리즈 모델의 미세 조정 버전으로, 모델 응답을 사용자 의도에 맞게 조정하는 인스트럭션 투닝의 힘을 보여 줍니다.



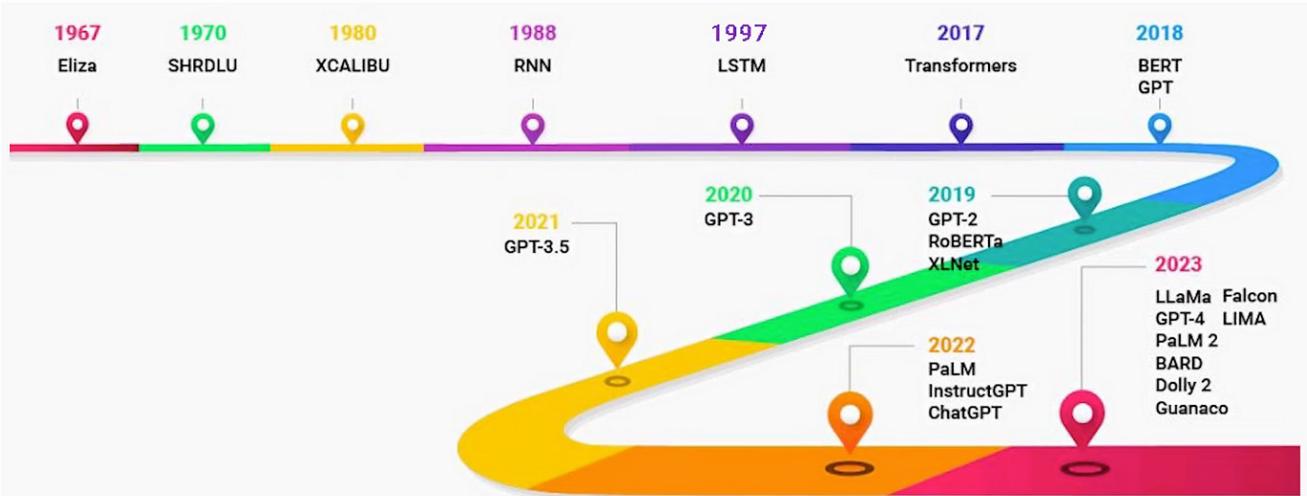
고급 대규모 언어 모델(LLM) 트레이닝 파이프라인에는 사전 트레이닝, 인스트럭션 투닝, 인간 피드백을 통한 강화 학습(RLHF) 또는 직접 선호도 최적화(DPO)를 사용한 선호도 정렬이 조합되어 있습니다.

## 7.4 기타 잘 알려진 대규모 언어 모델(LLM)

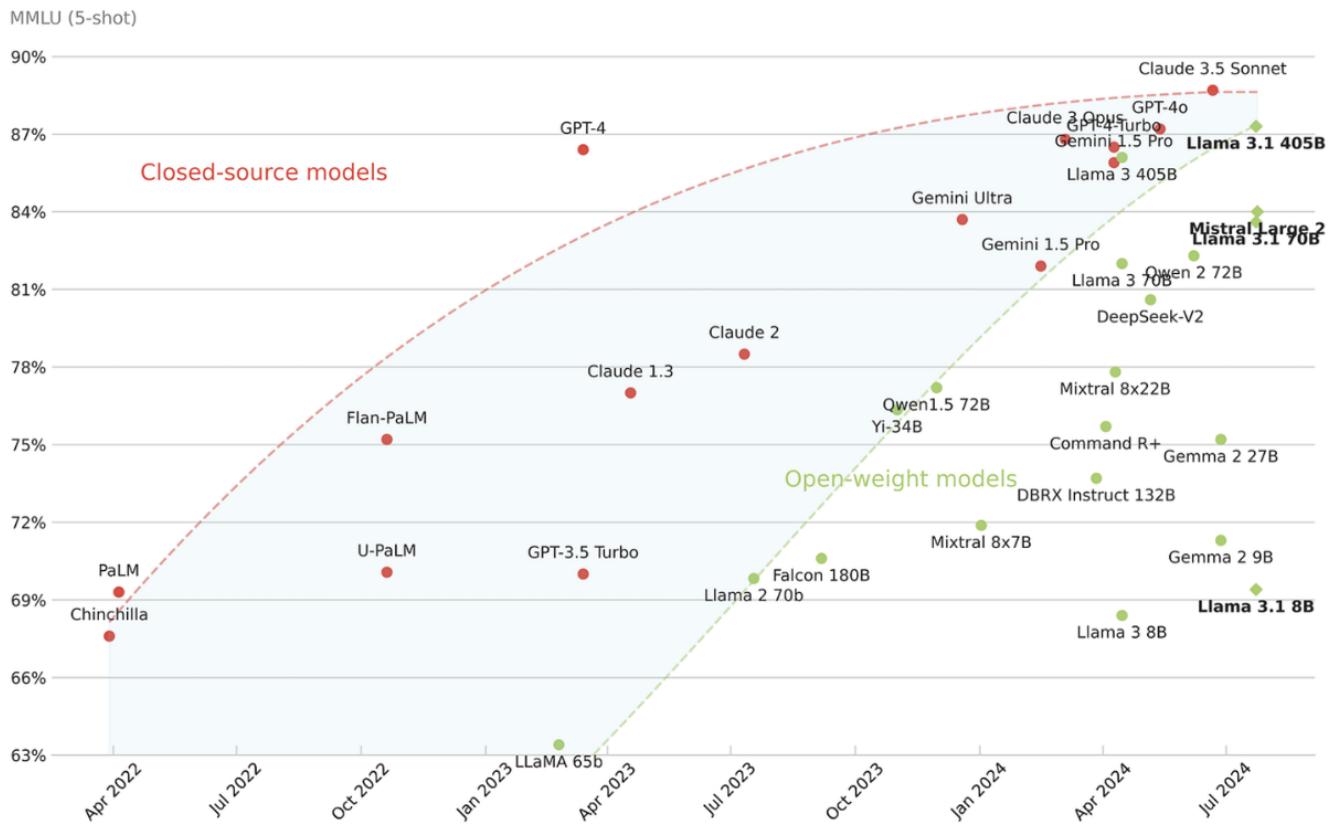
인공 지능의 고유한 기능과 발전을 제공하는 다양한 유명 모델들이 등장하면서 대규모 언어 모델(LLM)의 환경이 크게 풍성해졌습니다. 다음은 몇 가지 잘 알려진 LLM에 대한 업데이트된 개요입니다:

- **앤티로피의 클로드(2022)**: 인간의 가치에 부합하는 것을 목표로 AI 결과물에서 안전과 윤리적 고려를 우선시합니다.
- **Meta의 LLaMA(2023)**: 다양한 계산 요구에 맞는 다양한 크기의 모델을 제공하며 자연어 처리 벤치마크에서 인상적인 결과를 제공합니다.
- **Mistral.AI의 Mistral(2023)**: 오픈 소스 AI 솔루션에 중점을 두고 실시간 애플리케이션에 이상적인 고성능과 리소스 효율성의 균형을 맞추고 있습니다.
- **알리바바의 Qwen(2023)**: 영어와 중국어를 위한 고품질 이중 언어 AI 모델을 생성하여 다국어 애플리케이션을 촉진하고 혁신을 장려합니다.
- **Microsoft의 Phi(2023)**: 문맥 이해와 사용자 상호 작용을 위한 고급 훈련 기법을 통해 다양한 애플리케이션에서 다용도성과 통합을 강조합니다.
- **구글의 젬마 시리즈(2024년)**: 성능과 효율성에 중점을 둔 텍스트 생성, 요약, 추출 등 다양한 애플리케이션을 위한 경량의 최첨단 개방형 모델입니다.

## Evolution of Large Language Models



<https://www.analyticsvidhya.com/blog/2023/07/build-your-own-large-language-models/>



<https://medium.com/towards-data-science/fine-tune-llama-3-1-ultra-efficiently-with-unslot-7196c7165bab>

## 8. 멀티모달 모델(2023~현재)

### 8.1 GPT-4V(2023년) 및 GPT-4-o(2024년)

- **GPT-4V(2023)**는 이미 강력한 텍스트 기반 모델에 멀티모달 기능을 통합함으로써 AI 개발의 중요한 진전을 이루었습니다. 텍스트뿐만 아니라 이미지에서도 콘텐츠를 처리하고 생성할 수 있어 보다 포괄적인 AI 상호작용을 위한 토대를 마련할 수 있습니다.

---

## GPT-4 visual input example, Extreme Ironing:

---

User      What is unusual about this image?



Source: <https://www.barnorama.com/wp-content/uploads/2016/12/03-Confusing-Pictures.jpg>

---

GPT-4      The unusual thing about this image is that a man is ironing clothes on an ironing board attached to the roof of a moving taxi.

---

- GPT-4V에서 진화한 **GPT-4-o(2024)**는 정교한 문맥 이해와 함께 향상된 멀티모달 통합 기능을 제공합니다. 다양한 미디어에서 더 나은 일관성, 텍스트 프롬프트에서 고급 이미지 생성, 시각적 입력에 기반한 정교한 추론 기능을 제공함으로써 이전 버전 보다 개선되었습니다. 또한 GPT-4-o에는 윤리적 정렬을 위한 고급 훈련 메커니즘이 포함되어 있어 정확할 뿐만 아니라 책임감 있고 인간의 가치에 부합하는 결과물을 보장합니다.

GPT-4o 실시간 번역 라이브 데모.

## 8.2 Google의 쌍둥이자리(2023~현재)

- **제미니 프로(2023):** Google의 Gemini는 텍스트, 이미지, 오디오 및 비디오 처리를 통합하여 멀티모달 작업을 위해 설계된 모델 제품군을 소개합니다. 특히 Gemini Pro는 확장성과 효율성이 뛰어나 실시간 분석부터 다양한 미디어 형식의 복잡한 콘텐츠 생성에 이르기까지 다양한 애플리케이션에 고급 AI를 활용할 수 있습니다.
- **Gemini의 멀티모달 기능:** 다양한 규모의 애플리케이션을 위한 올트라 및 나노 버전을 포함한 Gemini 모델은 여러 데이터 유형에 대한 이해가 필요한 작업을 수행하도록 설계되었습니다. 동영상 요약, 다중 모드 번역, 대화형 학습 환경과 같은 작업에서 탁월한 성능을 발휘하며 멀티미디어 환경에서 AI의 역할을 발전시키려는 Google의 노력을 보여줍니다.

멀티모달 AI의 기능 | Gemini 데모

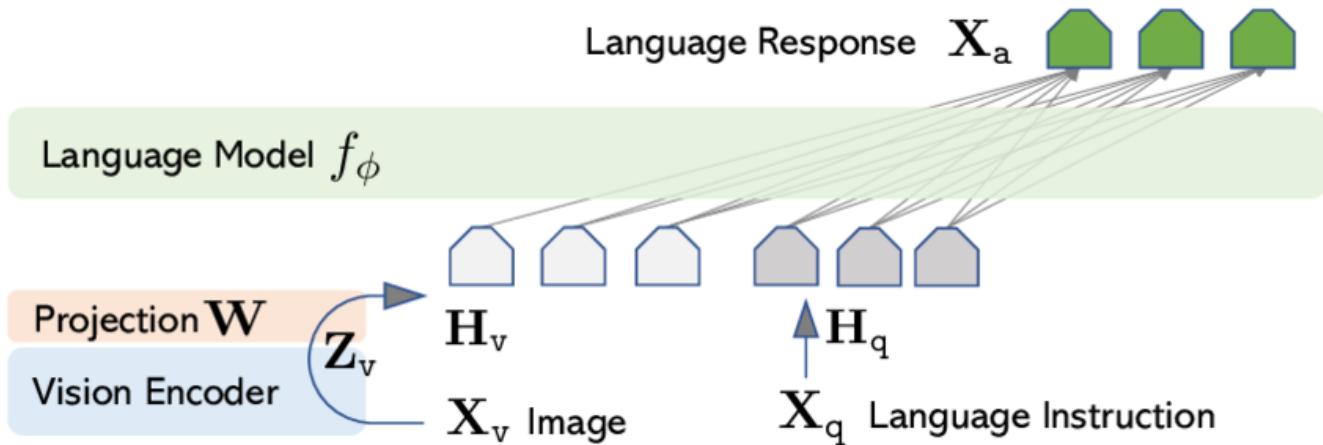
## 8.3 클로드 3.0 및 클로드 3.5(2023~현재)

- **2023년에 출시될 클로드 3.0**은 문맥 이해와 윤리적 고려 사항을 개선하여 AI 응답의 안전성과 신뢰성을 높이는 데 중점을 둔 모델입니다. 유해하거나 편향된 결과물을 피하기 위한 엄격한 준수를 유지하면서 보다 대화적이고 도움이 되도록 설계되었습니다.
- **Claude 3.5(2024)**는 Claude 3.0의 기능을 더욱 개선하여 복잡한 작업에서 더 나은 성능을 제공하고 처리 효율성을 높이며 사용자 요청을 더욱 미묘하게 처리할 수 있도록 합니다. 이 버전은 또한 멀티모달 상호작용을 강조하지만, 주로 텍스트 및 논리적 작업에 탁월하며 시각 또는 기타 감각 입력을 처리하는 새로운 기능을 통해 보다 통합된 사용자 경험을 제공합니다.

## 8.4 LLaVA(2023)

- **LLaVA(대규모 언어 및 시각 지원)**은 언어 이해와 시각 처리를 결합한 멀티모달 AI에 대한 혁신적인 접근 방식을 나타냅니다. 2023년에 개발된 LLaVA는 이미지를 해석하고 이를 텍스트 콘텐츠와 연관시켜 이미지에 대한 질문에 답하고 시각적 콘텐츠를

설명하거나 시각적 단서를 기반으로 텍스트를 생성할 수 있습니다. 이 아키텍처는 트랜스포머 모델의 강점을 활용하여 시각적 및 언어적 이해가 모두 필요한 작업에서 최첨단 성능을 달성합니다. 이 모델은 특히 오픈 소스 특성으로 인해 멀티모달 AI 애플리케이션에 대한 추가 연구 및 개발을 장려하는 것으로 유명합니다.



### LLaVA 아키텍처

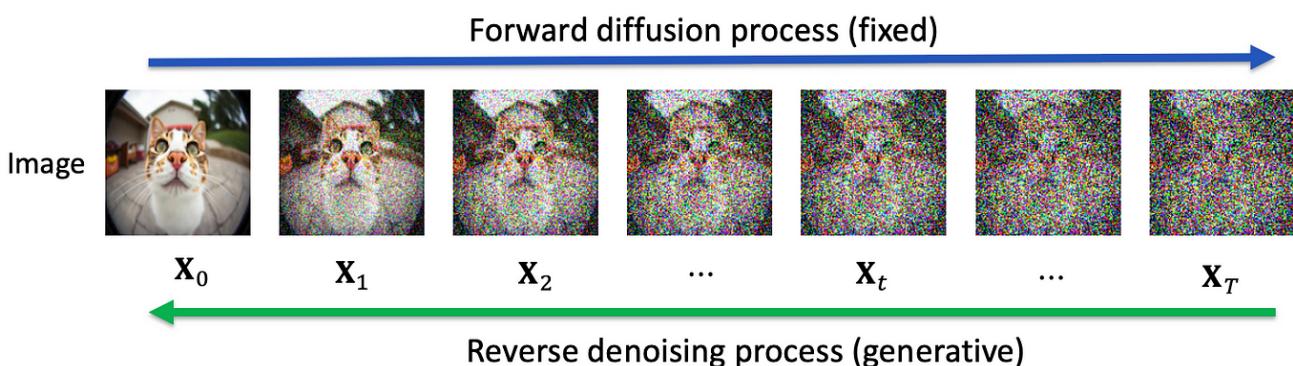
이러한 모델은 텍스트를 이해하고 생성할 뿐만 아니라 다양한 양식에 걸쳐 콘텐츠를 해석하고 생성하여 인간의 인지 능력을 더욱 가깝게 반영하는 AI 시스템으로의 전환을 의미합니다. 이러한 AI 모델의 진화는 보다 상호작용적이고 직관적이며 다양한 감각 입력이 혼합된 실제 시나리오를 처리할 수 있는 애플리케이션을 육성하여 일상 생활, 연구 및 산업 애플리케이션에서 AI가 달성할 수 있는 지평을 넓혀줍니다.

## 9. Diffusion Model 확산 모델(2015~현재)

확산 모델은 복잡한 데이터 분포에서 충실도가 높은 샘플을 생성하기 위한 새로운 방법론을 제공하면서 영향력 있는 생성 모델 범주로 부상하고 있습니다. 이러한 접근 방식은 수많은 애플리케이션에서 탁월한 성능을 발휘하는 점진적 노이즈 제거 기법을 사용함으로써 GAN 및 VAE와 같은 기존 모델과 대조를 이룹니다.

### 9.1 확산 모델 소개(2015)

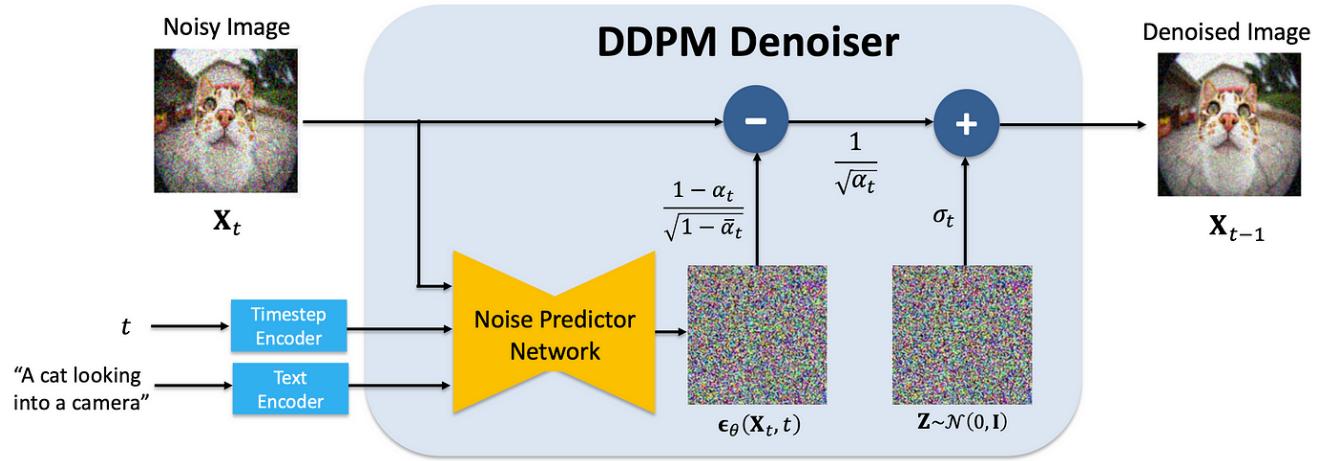
확산 모델을 소개하는 논문을 통해 Sohl-Dickstein 외(2015)가 그 토대를 마련했습니다. 이들은 점진적인 노이즈 추가를 역전시켜 노이즈를 다시 구조화된 데이터로 변환할 수 있는 생성 프로세스를 개념화했습니다.



### 9.2 확산 모델의 주요 특징

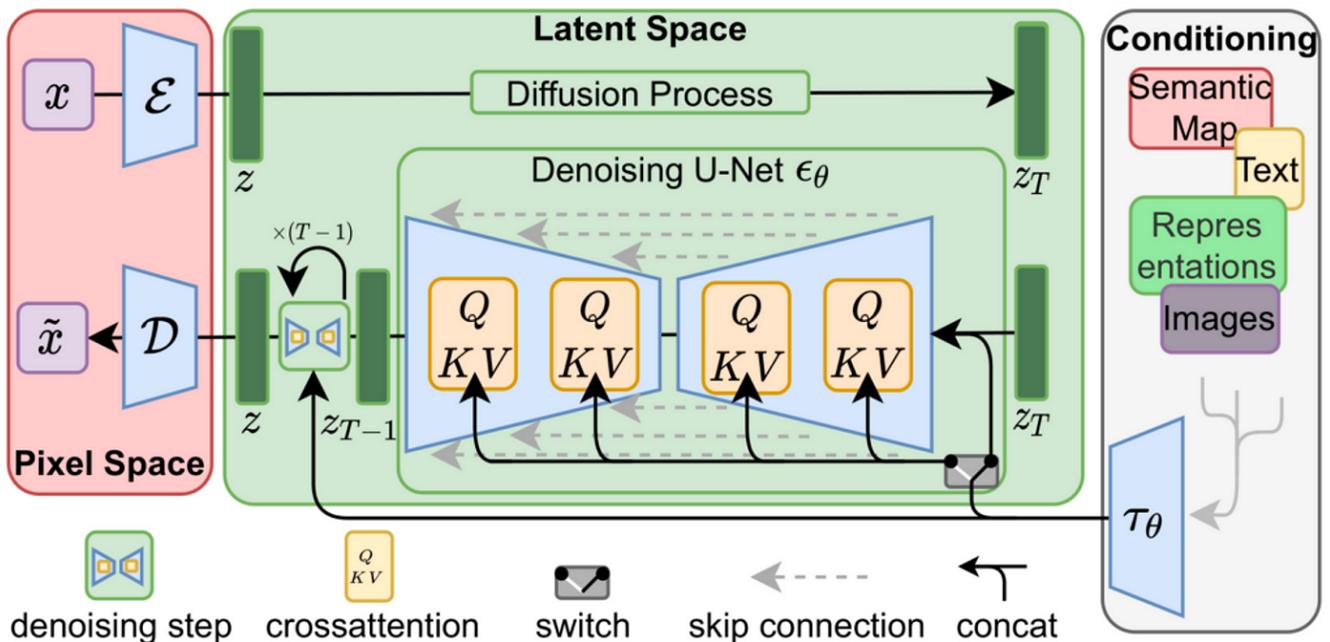
- **노이즈 제거 프로세스:** 이 모델은 단계적으로 노이즈를 추가하고(포워드 프로세스) 이를 역으로(백워드 프로세스) 학습하여 효과적으로 노이즈를 제거하여 샘플을 생성합니다.
- **마르코프 체인:** 두 프로세스 모두 마르코프 체인으로 구조화되어 있으며, 각 순방향 단계마다 가우스 노이즈가 추가되고, 모델은 이를 역으로 제거하는 방법을 학습합니다.
- **훈련 목표:** 목표는 각 단계에서 예측된 노이즈와 실제 노이즈의 차이를 최소화하여 증거 하한(ELBO)의 형태를 최적화하는 것입니다.

- 안정성과 견고성: GAN보다 안정성이 뛰어나 모드 붕괴와 같은 문제를 방지하여 다양하고 고품질의 결과물을 일관되게 생성합니다.



### 9.3 확산 모델의 발전(2020~현재)

- 노이즈 제거 확산 확률론적 모델(DDPM) (2020): 확산 프로세스를 개선하여 이미지 합성의 새로운 벤치마크를 설정했습니다.
- 노이즈 제거 확산 암시적 모델(DDIM)(2021): 향상된 비 마르코비안 샘플링으로 효율성이 향상되어 생성 프로세스가 더욱 유연해졌습니다.
- 확률 미분  
방정식을 통한 점수 기반 생성 모델 (2021): 효율적인 샘플 생성을 위해 확률 미분 방정식 활용.
- 잠재 확산 모델(2022): 안정적 확산과 같이 널리 사용되는 텍스트-이미지 생성 시스템의 기반이 되어 AI 생성 이미지 분야를 크게 발전시키고 보다 접근하기 쉽고 효율적인 생성 AI 도구의 길을 열었습니다.



잠재 확산 모델의 아키텍처

### 9.3 텍스트-이미지 생성

DALL-E 3 및 Stable Diffusion 3와 같은 모델은 텍스트 설명에서 고품질 이미지를 생성하는 데 탁월하며, DALL-E 3는 상세하고 정확한 비주얼을 제공하고 Stable Diffusion은 이미지 생성 기술에 대한 액세스를 대중화하는 오픈 소스 대안을 제공합니다.

DALL-E 3



MIDJOURNEY 5.2



STABLE XL



<https://generativeai.pub/dall-e-3-vs-midjourney-5-2-vs-stable-xl-same-prompt-different-resultsa68ae19b223e>

- **FLUX.1 (2024):** 블랙 포레스트 랩은 뛰어난 속도와 품질, 신속한 적용을 제공하는 AI 이미지 생성을 위한 고급 확산 모델인 FLUX.1을 공개했습니다. 슈넬, 개발, 프로의 세 가지 버전으로 제공되는 FLUX.1은 정류된 흐름 트랜스포머와 같은 혁신적인 기술을 활용하여 매우 사실적인 이미지를 생성합니다. FLUX.1은 텍스트를 생성하고 손가락과 발가락과 같은 디테일을 처리하는 등 훌륭한 이미지 생성기에 필요한 모든 기능을 제공합니다.



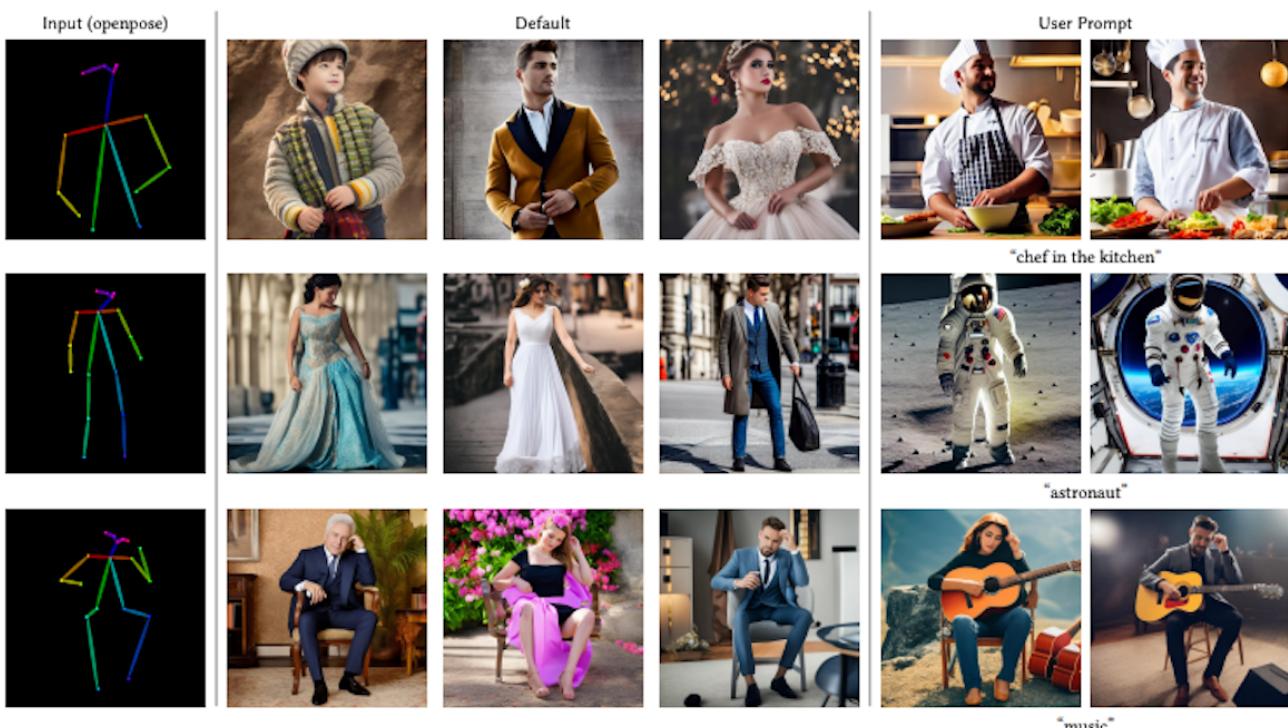
"측면에 FLUX.1이라고 적힌 커피잔"이라는 간단한 문구가 있는 FLUX.1 Shenell 모델에서 생성된 이미지. "FLUX.1"이라는 텍스트가 선명하게 보이는 커피잔의 고품질 이미지로, FLUX.1의 텍스트 생성 기능을 보여줍니다.(출처)

- **드림부스(2022):** 특정 피사체의 몇 장의 이미지로 확산 모델을 훈련하여 개인화된 이미지를 생성할 수 있습니다.
- **LoRA(2022):** 최소한의 추가 매개변수로 확산 모델을 미세 조정하여 특정 작업이나 데이터 세트에 모델을 더 쉽게 적용할 수 있는 기술인 로우랭크 적용(Low-Rank Adaptation)의 약자입니다.



단일 개념 생성의 질적 비교. 각 컨셉에 대한 참조 이미지가 왼쪽 열에 나와 있습니다. LoRA 기반 방식은 충실도 측면에서 커스텀 디퓨전보다 성능이 뛰어납니다. 또한 직교 적응과 SBoRA는 믹스 오브 쇼와 비슷한 성능을 보이지만 직교 제약 조건을 도입하여 다중 콘셉트 시나리오에서 이점을 제공합니다.

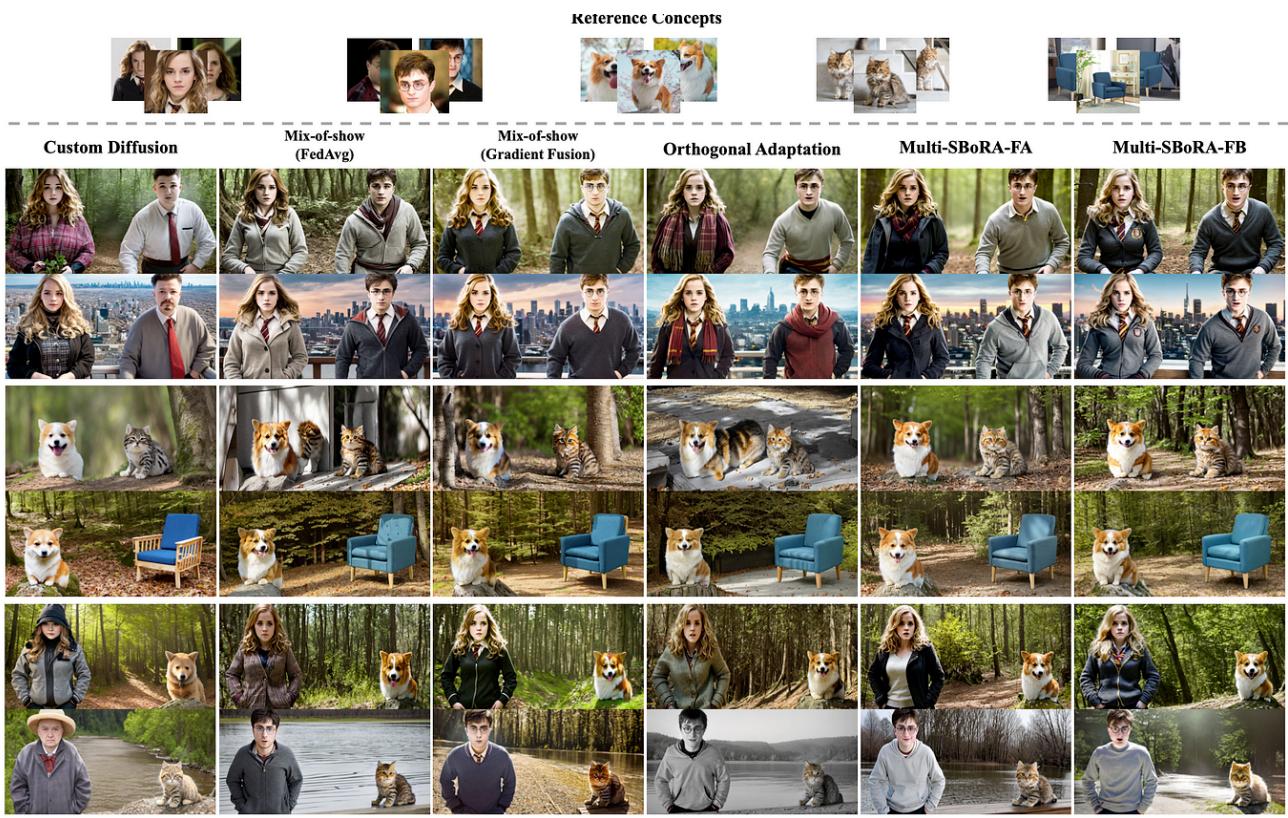
- **ControlNet (2023):** 스케치나 데스 맵과 같은 추가 입력에 따라 확산 모델을 조절하여 생성된 이미지를 더 잘 제어할 수 있습니다.



포즈 제어 기능이 있는 ControlNet을 통한 안정적인 확산([출처](#))

멀티 컨트롤넷을 사용한 애니디프 | StableDiffusion.

- **Multi-SBoRA(2024):** Multi-SBoRA는 여러 개념에 맞게 확산 모델을 커스터마이징하는 새로운 방법입니다. 직교 표준 기저 벡터를 사용하여 미세 조정을 위한 낮은 순위의 행렬을 구성함으로써 개념 간 간섭을 줄이는 지역 및 비중첩 가중치 업데이트가 가능합니다. 이 접근 방식은 사전 학습된 모델의 지식을 보존하고 계산 오버헤드를 줄이며 모델 유연성을 향상시킵니다. 실험 결과에 따르면 Multi-SBoRA는 다중 개념 사용자 지정에서 최적의 성능을 달성하는 동시에 독립성을 유지하고 크로스톡 효과를 완화하는 것으로 나타났습니다.



다중 개념 생성의 질적 비교. 결과는 (1) 캐릭터 생성, (2) 개체 생성, (3) 캐릭터와 개체 결합 생성의 세 가지 경우로 구분됩니다. 각 개념에 대한 참조 이미지는 맨 윗줄에 나와 있습니다. 커스텀 디퓨전 및 믹스 오브 쇼와 같이 직교 디자인이 없는 방법은 특히 복잡한 얼굴 특징을 가진 캐릭터에서 컨셉의 정체성이 크게 손실되는 것을 볼 수 있습니다. 직교 적용은 아이덴티티 보존이 개선되었지만 모델의 전반적인 지식이 손상되어 봉고로 이어질 수 있습니다. 반면에 우리가 제안한 방법은 각 개념의 아이덴티티를 효과적으로 보존하면서 보다 안정적인 생성을 보장하는 우수한 결과를 달성합니다.

확산 모델 연구의 궤적은 속도와 품질을 최적화하면서 다양한 AI 아키텍처의 강점을 결합하는 통합 모델의 잠재력과 함께 유망한 미래를 보여줍니다.

## 9.4 텍스트-비디오: OpenAI 소라(2024)

OpenAI Sora는 OpenAI의 멀티모달 AI 제품의 기능을 확장하는 새로운 텍스트-비디오 생성 모델입니다. 이 모델을 통해 사용자는 텍스트 설명에서 동영상을 생성하여 텍스트와 동적 시각 콘텐츠 사이의 간극을 효과적으로 메울 수 있습니다. 멀티모달 프레임워크에 소라를 통합하면 창의적인 애플리케이션의 잠재력이 향상되어 사용자는 최소한의 입력으로 풍부한 멀티미디어 콘텐츠를 생성할 수 있습니다. 이러한 발전은 복잡한 형태의 미디어를 이해하고 생성할 수 있는 보다 직관적이고 대화형 AI 시스템을 위한 중요한 진전을 의미합니다.

OpenAI 소라가 작동 중입니다: 도쿄 산책

## 10. 결론

AI와 딥러닝의 역사는 상당한 발전과 변화의 혁신으로 점철되어 있습니다. 초기 신경망부터 컨볼루션 신경망(CNN), 순환 신경망(RNN), 트랜스포머, 확산 모델과 같은 정교한 아키텍처에 이르기까지 이 분야는 다양한 영역에 혁명을 일으켰습니다.

최근의 발전으로 OpenAI의 GPT-4o, Google의 Gemini Pro, Anthropic의 Claude 3.5 Sonnet, Meta의 LLaMA3.1과 같이 뛰어난 자연어 및 다중 모달리티 기능을 보여주는 대규모 언어 모델(LLM)과 대규모 다중 모달 모델(LMM)이 개발되고 있습니다. 또한 미드저니, DALL-E 3, 스테이블 디퓨전, FLUX.1, 소라와 같은 텍스트-이미지 및 텍스트-비디오 생성 모델을 포함한 제너레이티브 AI의 획기적인 발전은 AI의 창의적인 잠재력을 확장했습니다.

확산 모델은 또한 다양한 응용 분야를 가진 강력한 생성 모델로 부상했습니다. 보다 효율적이고 해석 가능하며 유능한 모델을 개발하기 위한 연구가 계속 진행됨에 따라 AI와 딥러닝이 사회와 기술에 미치는 영향은 더욱 커질 것입니다. 이러한 발전은 전통적인 분야에서 혁신을 주도하고 창의적 표현, 문제 해결, 인간과 AI의 협업을 위한 새로운 가능성을 창출하고 있습니다.

하지만 딥러닝이 AI에 대한 유일한 접근 방식이거나 최선의 접근 방식은 아닙니다. 기호적 AI, 강화 학습, 신경 기호적 AI는 고유한 강점을 제공하며 해석 가능성 및 계산 리소스 요구 사항과 같은 딥러닝의 한계를 해결합니다. AI에 대한 포괄적인 관점에는 이러한 다양한 방법론이 포함되어야 합니다.

AI의 미래는 다양한 접근 방식의 시너지에 달려 있습니다. 연구가 발전함에 따라 다양한 AI 기술 생태계를 육성하면 균형 잡힌 효과적인 진화가 이루어지고 사회와 기술 모두에게 이익이 될 것입니다.