

# 강의계획표

주	해당 장	주제
1	1장	머신러닝이란
2	2장, 3장	머신러닝을 위한 기초지식, 구현을 위한 도구
3	4장	선형 회귀로 이해하는 지도학습
4	5장	분류와 군집화로 이해하는 지도 학습과 비지도 학습
5	6장	다양한 머신러닝 기법들 - 다항 회귀, Logistic Regression - 정보이론, 결정트리 - SVM, Ensemble
6		
7		
8		중간고사 (04-20)
9	7장	인공 신경망 기초 - 문제와 돌파구
10	8장	고급 인공 신경망 구현
11	9장	신경망 부흥의 시작, 합성곱 신경망
12	10장	순환 신경망
13	11장	차원축소와 매니폴드 학습
14	12장	오토인코더와 잠재표현 학습
15	<b>13장</b>	<b>AI의 현재와 미래</b>
16		기말고사

# 13장 인공지능의 현재와 미래

## 14 주차

- 기계학습의 여러 기법 정리
- 실시간 객체탐지에 활용하고 있는 YOLO
- 어텐션, 트랜스포머, BERT, GPT-3, T5
- 적대적 생성모델, GAN
- 강화학습
- 인공지능의 윤리적 문제
- 참고사이트
  - [YOLO, Object Detection Network - 블로그 - 네이버](#)
  - [\[딥러닝\] Object Detection with YOLO - velog](#)
  - [딥 러닝을 이용한 자연어 처리 입문](#)
  - [GPT3, BERT까지! 자연어처리 알고리즘발전 스토리 \[토크아이티\] ...](#)
  - [\[GAN\] 이제 제발 이해하자 GAN - 0장 \(개념\) : 네이버 블로그](#)
  - [강화학습 Reinforcement Learning – 생활코딩](#)
  - [\[NHN FORWARD 2020\] 강화 학습 기초: #1 강화 학습 이해하기](#)
  - [9.인공지능 윤리\(중급\) Part 2. AI 윤리 문제 사례 - AI4School](#)

# 인공 신경망으로 무엇을 할 수 있을까? (1)

## ❖ 학습내용



# 인공 신경망으로 무엇을 할 수 있을까? (2)

## ❖ CNN 활용

- 시각 지능 구현
- 객체위치 파악 localization
- 실시간 객체탐지 object detection YOLO
- 비디오와 텍스트 처리 video and text processing
- Image 생성

## ❖ RNN 활용

- 언어 지능 구현: 챗봇, 기계번역, 질의응답, 영상캡션 추가, 내용요약, MRC
- 시퀀스-투-시퀀스 sequence-to-sequence: seq2seq
- 어텐션 메커니즘 attention mechanism
- 트랜스포머 transformer
- GPT generative pretrained transformer\_3

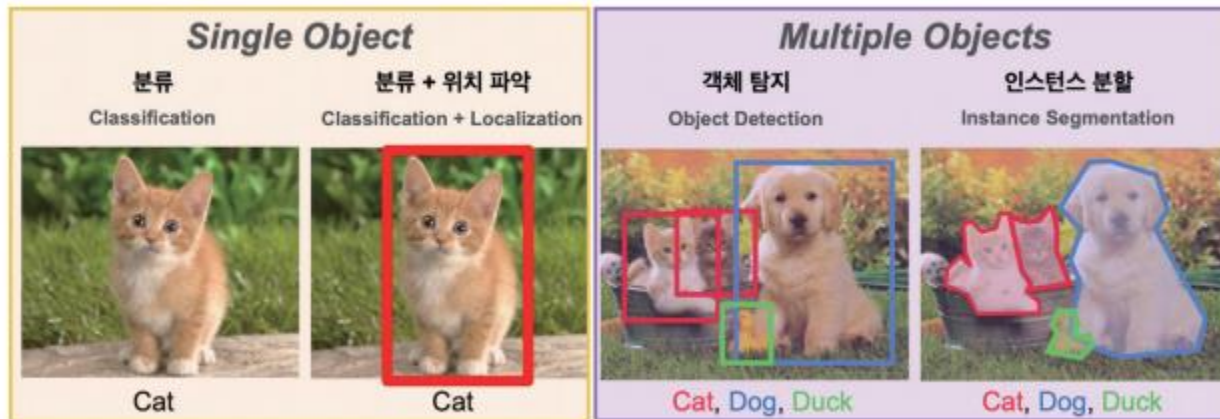
## ❖ 적대적 생성 모델: GAN

## ❖ 강화학습

# 시각 인식 응용 (1)

## ❖ 실시간 객체탐지 기술

- 경계 상자 bounding box
- 객체탐지 object detection : 하나의 이미지에서 여러 문체를 분류하고 위치 추정
- 이미지 분류 image classification
- 다중-레이블 분류 multi-labeled classification
- 인스턴스 분할 instance segmentation : 인스턴스별로 레이블을 구분



# 시각 인식 응용(2)

## ❖ YOLO You Only Look Once

- 2015년 조셉 레드몬 Joseph Redmon
- <https://homl.info/yolodemo>
- 한번 보고 처리: 이미지 전체를 단 한번만 봄
- 모델 사용: 단 하나의 인공 신경망을 통해 처리
- 실시간 객체 탐지: 실시간으로 여러 장의 이미지를 탐지
- [YOLO, Object Detection Network - 블로그 - 네이버](#)
- [\[딥러닝\] Object Detection with YOLO - vlog](#)

## ❖ 시맨틱 분할 semantic segmentation

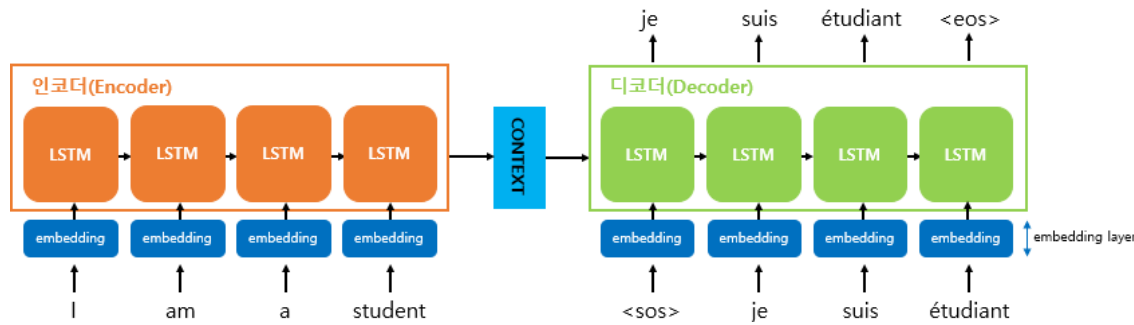
- 이미지 내의 모든 픽셀에 대해 어떤 클래스 레이블에 속하는지 예측



# 언어처리 기술 응용 (1)

## ❖ 시퀀스-투-시퀀스 seq2seq

- 챗봇, 기계번역, 질의응답, 영상캡션 추가, 내용요약, STT 응용



I	0.157	am	0.78
	-0.25		0.29
	0.478		-0.96
	-0.78		0.52
a	0.75	student	0.88
	-0.81		-0.17
	0.96		0.29
	0.12		0.48

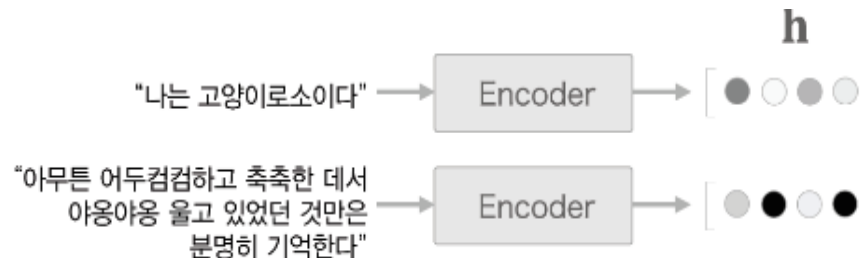
## ■ 컨텍스트 벡터 context vector:

- 입력 시퀀스를 고정된 크기의 벡터 표현으로 압축
- 인코더에서의 마지막 RNN 셀의 은닉 상태값

CONTEXT	0.15
	0.21
	-0.11
	0.91

## ■ 문제점

- 인코더가 고정된 크기의 벡터에 정보를 압축하기 때문에 정보 손실 발생
- 순환 신경망의 고질적인 문제인 기울기 소실 문제(장기 의존성) 발생





# 언어처리 기술 응용 (2)

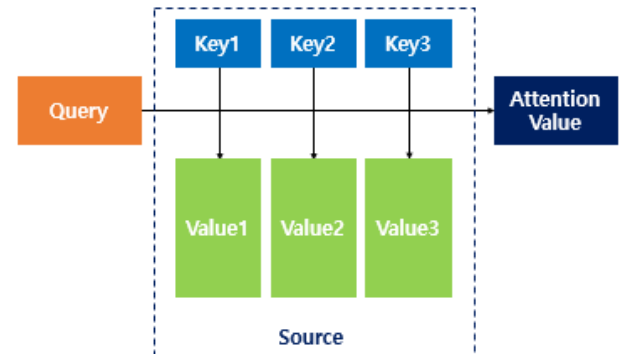
## ❖ 어텐션attention

- 디코더가 출력 단어를 예측할 때 시점(time step)마다, 인코더에서 전체 입력 문장을 다시 한번 참고함
- 해당 시점에서 예측해야 할 단어와 연관이 있는 단어에 가중치를 더 부여함으로써 중요한 부분에 집중

- 어텐션 함수

- $\text{Attention}(Q, K, V) = \text{Attention Value}$

Q = Query : t 시점의 디코더 셀에서의 은닉 상태  
K = Keys : 모든 시점의 인코더 셀의 은닉 상태들  
V = Values : 모든 시점의 인코더 셀의 은닉 상태들



- 주어진 '쿼리(Query)'에 대해서 모든 '키(Key)'와의 유사도를 각각 구하고,
- 이 유사도를 키와 맵핑되어 있는 각각의 '값(Value)'에 반영한 후,
- 유사도가 반영된 '값(Value)'을 모두 더해서 리턴

# 언어처리 기술 응용 (3)

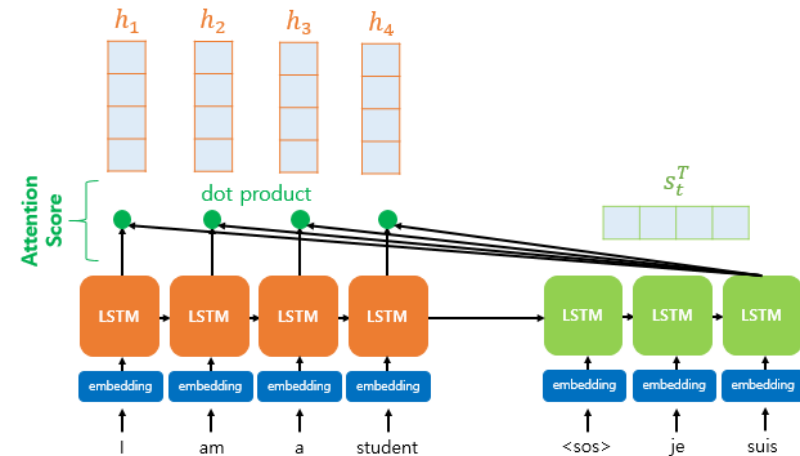
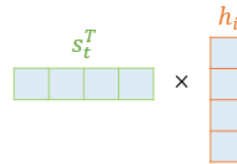
## ❖ Dot-Product Attention: Luong 어텐션

### 1) 어텐션 스코어(Attention Score)

- 현재 디코더의 시점  $t$ 에서 단어를 예측하기 위해, 인코더의 모든 은닉 상태( $h_1, h_2, \dots, h_N$ ) 각각이 디코더의 현재 시점의 은닉 상태  $s_t$  와 얼마나 유사한지를 판단하는 스코어값

$$\text{score}(s_t, h_i) = s_t^T h_i$$

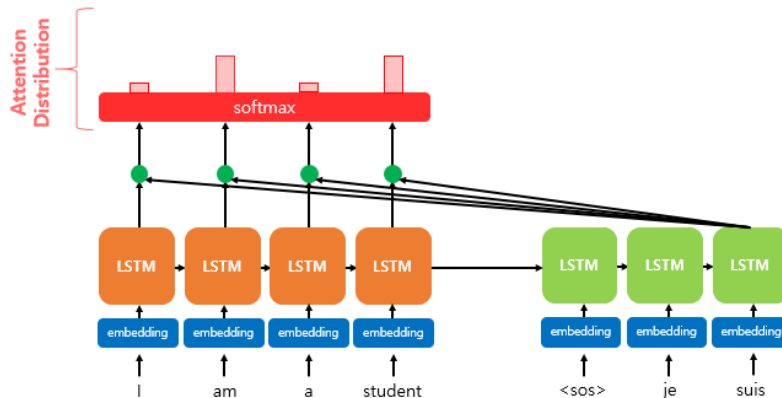
$$e^t = [s_t^T h_1, \dots, s_t^T h_N]$$



### 2) 소프트맥스함수를 통해 어텐션 분포

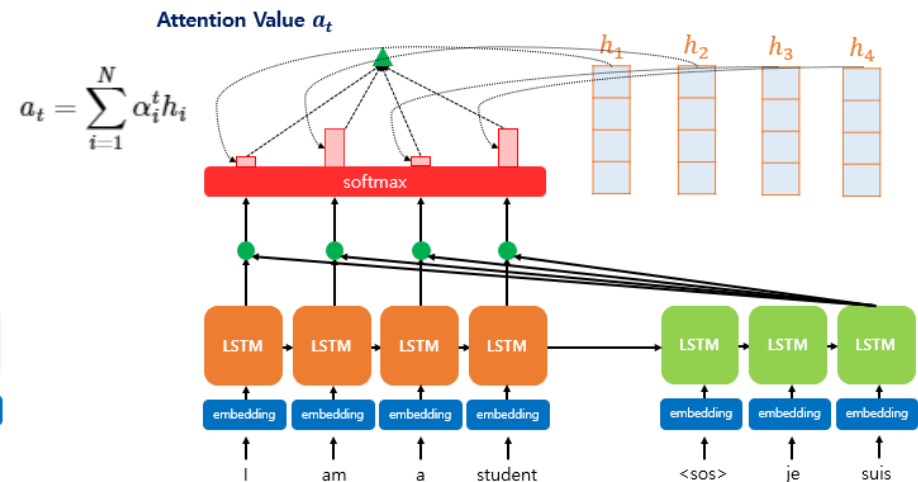
- 어텐션 가중치(Attention Weight)

$$\alpha^t = \text{softmax}(e^t)$$



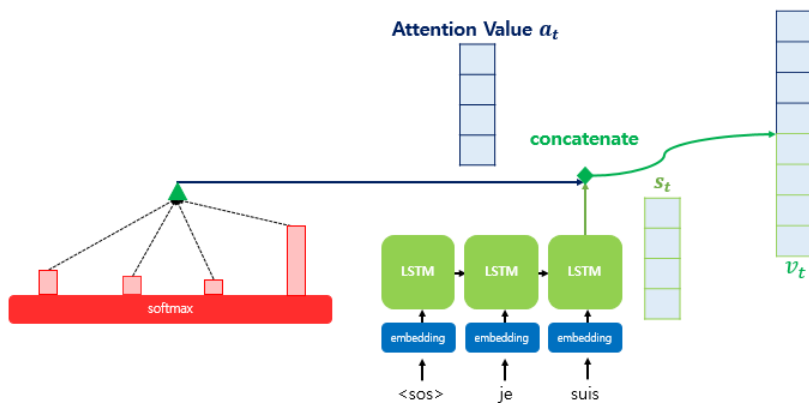
### 3) 어텐션 값(Attention Value)

- 각 인코더의 어텐션 가중치와 은닉 상태를 가중합



# 언어처리 기술 응용 (4)

## 4) 어텐션 값과 디코더의 t 시점의 은닉 상태를 연결



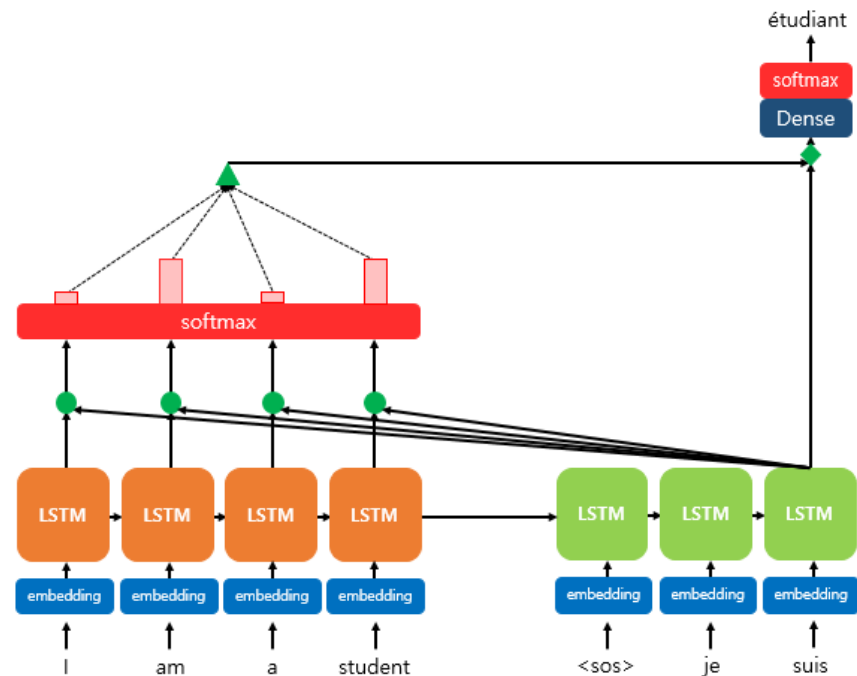
## 5) 출력층 연산의 입력이 되는 $\tilde{s}_t$ 를 계산

$$\tanh \left( \begin{matrix} \text{Grid} \\ W_c \end{matrix} \times \begin{matrix} \text{Vector} \\ v_t \end{matrix} \right) = \begin{matrix} \text{Vector} \\ \tilde{s}_t \end{matrix}$$

$$\tilde{s}_t = \tanh(W_c[a_t; s_t] + b_c)$$

## 6) $\tilde{s}_t$ 를 출력층의 입력으로 사용

$$\hat{y}_t = \text{Softmax}(W_y \tilde{s}_t + b_y)$$



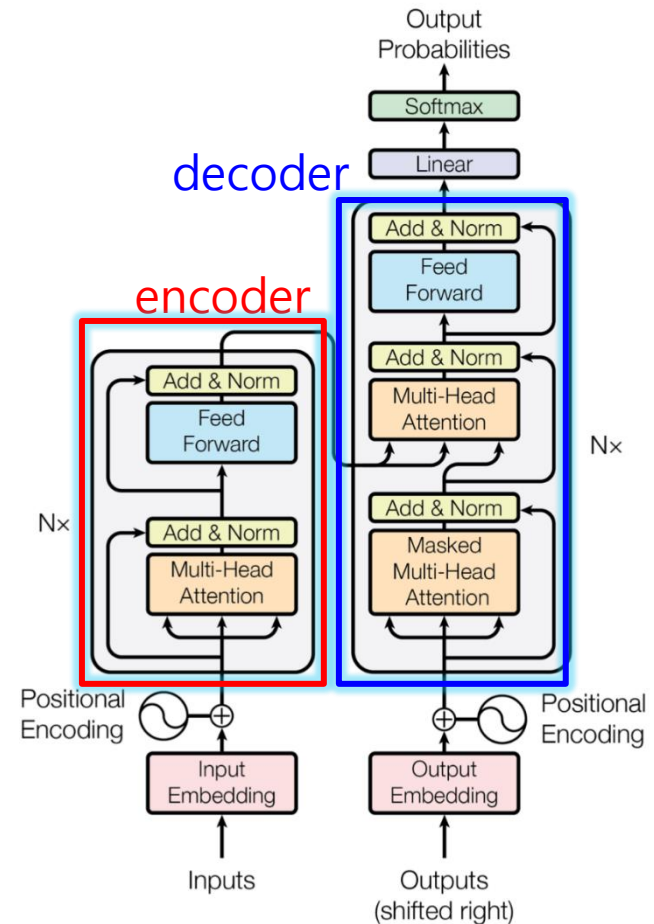
# 언어처리 기술 응용 (5)

## ❖ 트랜스포머 Transformer

- 2017년 구글 논문, "Attention is all you need"
- 어텐션만으로 인코더와 디코더를 구현한 모델
- Self Attention
- Multi-head attention
- Masked Language Model

## ❖ BERT

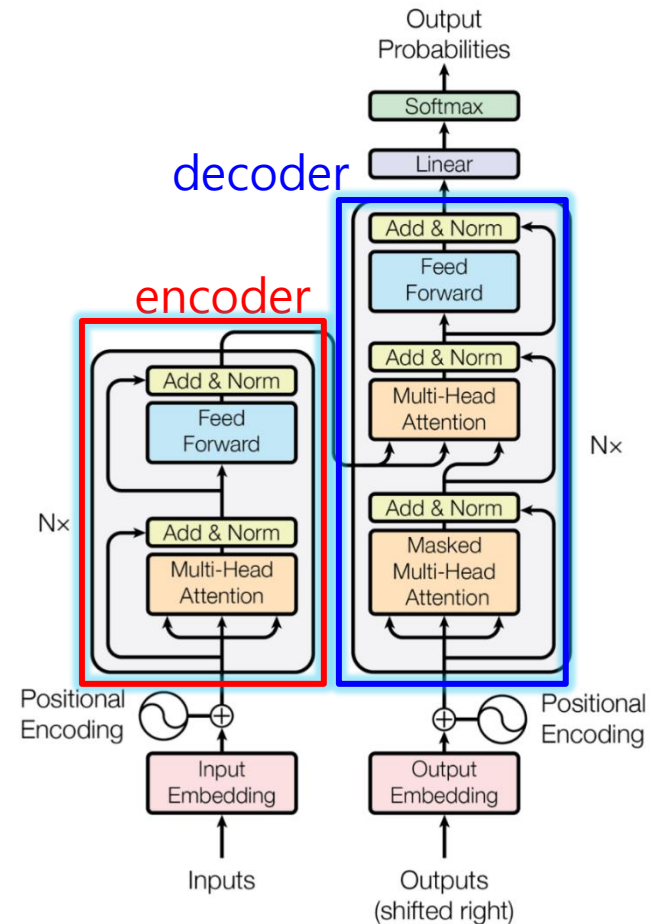
- Bidirectional Encoder Representations from Transformers
- 위키피디아(25억 단어)와 BooksCorpus(8억 단어)와 같은 레이블이 없는 텍스트 데이터로 사전 훈련된(Pre-Trained) 언어 모델
- Masked Language Model
- 다음 문장 예측(Next Sentence Prediction, NSP)
- Fine-Tuning



# 언어처리 기술 응용 (6)

## ❖ GPT Generative Pre-trained Transformer -3

- [OpenAI API](#)
- 1,750억개의 매개변수, 3,000억개의 text
- 문제풀이, 랜덤 글짓기, 간단한 사칙연산, 번역, 주어진 문장에 따른 간단한 웹 코딩
- chatGPT
- BARD
- [GPT-3 패러다임을 바꿀 미친 성능의 인공지능 등장 및 활용 사례 ...](#)
- [NAVER AI NOW](#)
- [새로운 AI의 시작, HyperCLOVA](#)



# 적대적 생성 모델: GAN

## ❖ 생성적 모델링 generative modeling

- 학습 데이터를 이용한 훈련을 통해 데이터의 분포를 학습
- 랜덤 노이즈를 학습을 통해 익힌 분포와 일치하도록 만드는 생성 모델
- 데이터 샘플링을 통해, 기존에 관찰했던 데이터와 같은 분포를 갖지만 존재하지는 않는 새로운 데이터를 생성



## ❖ GAN(Generative Adversarial Network)

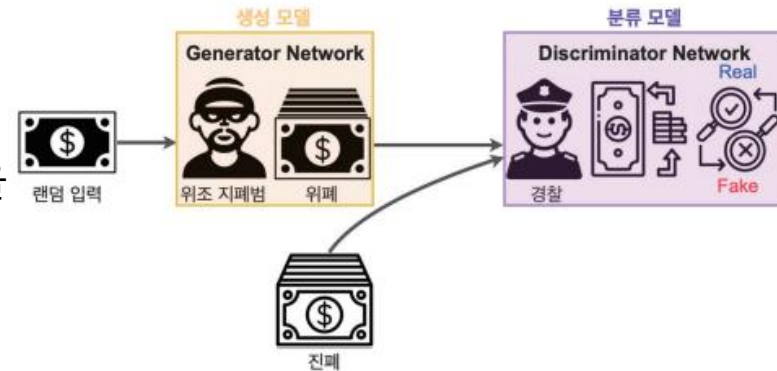
- 2014년 **이안 굿펠로** Ian Goodfellow
- 분류 문제를 해결하는 **판별자 네트워크** discriminative networks 와 **생성자 네트워크** generator networks 의 대립을 통해서 진짜 같은 가짜 이미지를 만듦

## ❖ 응용

- 예술, 패션 및 광고, 비디오 게임, 영화 산업
- **딥페이크** deep fake

## ❖ 참고사이트

- [\[GAN\] 이제 제발 이해하자 GAN - 0장 \(개념\) : 네이버 블로그](#)
- [Deep Dream Generator](#)



# 강화학습 (1)

## ❖ 강화학습 reinforcement learning

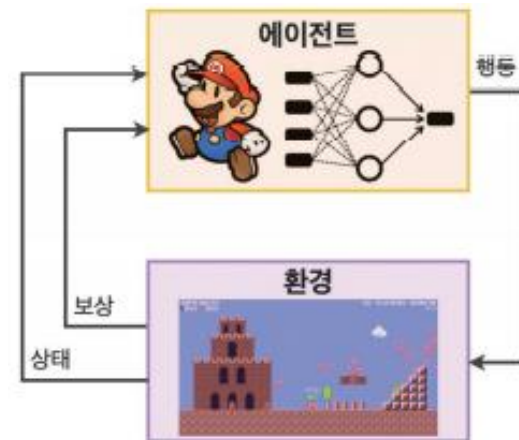
- 행동심리학
- 어떤 환경 안에서 정의된 에이전트가 현재의 상태를 인식하여, 선택 가능한 행동들 중 보상reward를 최대화하는 행동 혹은 행동 순서를 선택하는 방법
- 에이전트가 학습을 하는 과정에서 점점 발전하게 되는 의사결정 전략을 정책policy

## ❖ 마르코프 의사결정 프로세스 Markov decision process

- 상태state: 정적인 요소 + 동적인 요소
- 행동action: 어떠한 상태에서 취할 수 있는 행동
- 보상reward: 에이전트가 학습할 수 있는 유일한 정보
- 정책policy: 순차적 행동 결정문제에서 구해야 할 답, 모든 상태에 대해 에이전트가 어떤 행동을 해야 하는지 정해 놓은 것

## ❖ 참고사이트

- [강화학습 Reinforcement Learning – 생활코딩](#)
- [\[NHN FORWARD 2020\] 강화 학습 기초: #1 강화 학습 이해하기](#)



# 강화학습 (2)

## ❖ AlphaGo 발전

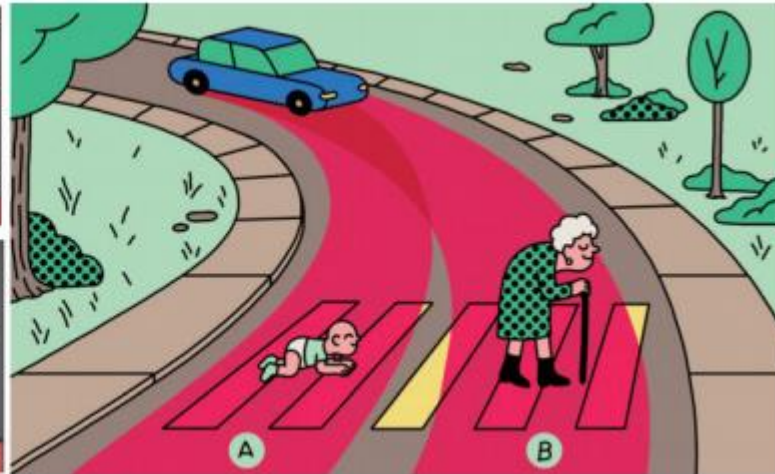
### ▪ 알파고 - 나무위키





# 인공지능과 윤리적 딜레마: 윤리적 기계 (1)

- 데이터와 알고리즘 기반으로 동작하는 인공지능의 판단이 과연 객관적이고 공정한 것일까?
- 인공지능의 편향성 문제와 윤리 문제가 대두
- 미국 뉴욕대 AI 나우 연구소는 범죄예측시스템 운용 경험이 있는 미국 13개 시 경찰 중 9곳에서 인종이나 성적 차별에 근거한 편견과 오류가 발견
- 흑인이 백인에 비해 범죄의 용의자일 확률이 높다고 표시하는 범죄 예측 시스템의 치명적인 **편향**bias
- 인공지능(AI) 챗봇 '이루다' 서비스: 성소수자 혐오, 개인정보 유출



# 인공지능과 윤리적 딜레마: 윤리적 기계 (2)

## ❖ AI의 편향bias

- **인간의 편향human bias**: 학습하는 데이터는 편향을 가질 수 있는 인간이 만듦
- **숨겨진 편향hidden bias**: 잘 드러나지도 찾을 수도 없는 가장 개선하기 어려운 편향
- **데이터 표본 편향data sampling bias**: 자료 수집 단계의 문제로 인한 데이터 샘플링 편향으로 예를 들어 데이터 중 남자가 의사인 경우가 많고 여자는 간호사가 많은 경우. 남자는 의사, 여자는 간호사로 단정하는 경우
- **롱테일 편향long-tail bias**: 학습 데이터에서 특정 종류의 데이터가 빠져 생기는 편향으로 자율주행 AI 시스템 개발에 큰 걸림돌
- **고의적 편향**: 해킹이나 공격으로 인해 AI에 의도적으로 편향을 일으킬 가능성

# 인공지능과 윤리적 딜레마: 윤리적 기계 (3)

## ❖ 인공지능 윤리 (참고사이트)

- [1. 윤리적 인공지능](#)
- [2. 윤리적 인공지능 - 책임성](#)
- [3. 윤리적 인공지능 - 공정성](#)
- [4. 윤리적 인공지능 - 투명성](#)
- [5. 윤리적 인공지능 - 안전성](#)
- [6. AI 윤리적 대응](#)