

지능형 디지털금융과 신뢰가능한 인공지능

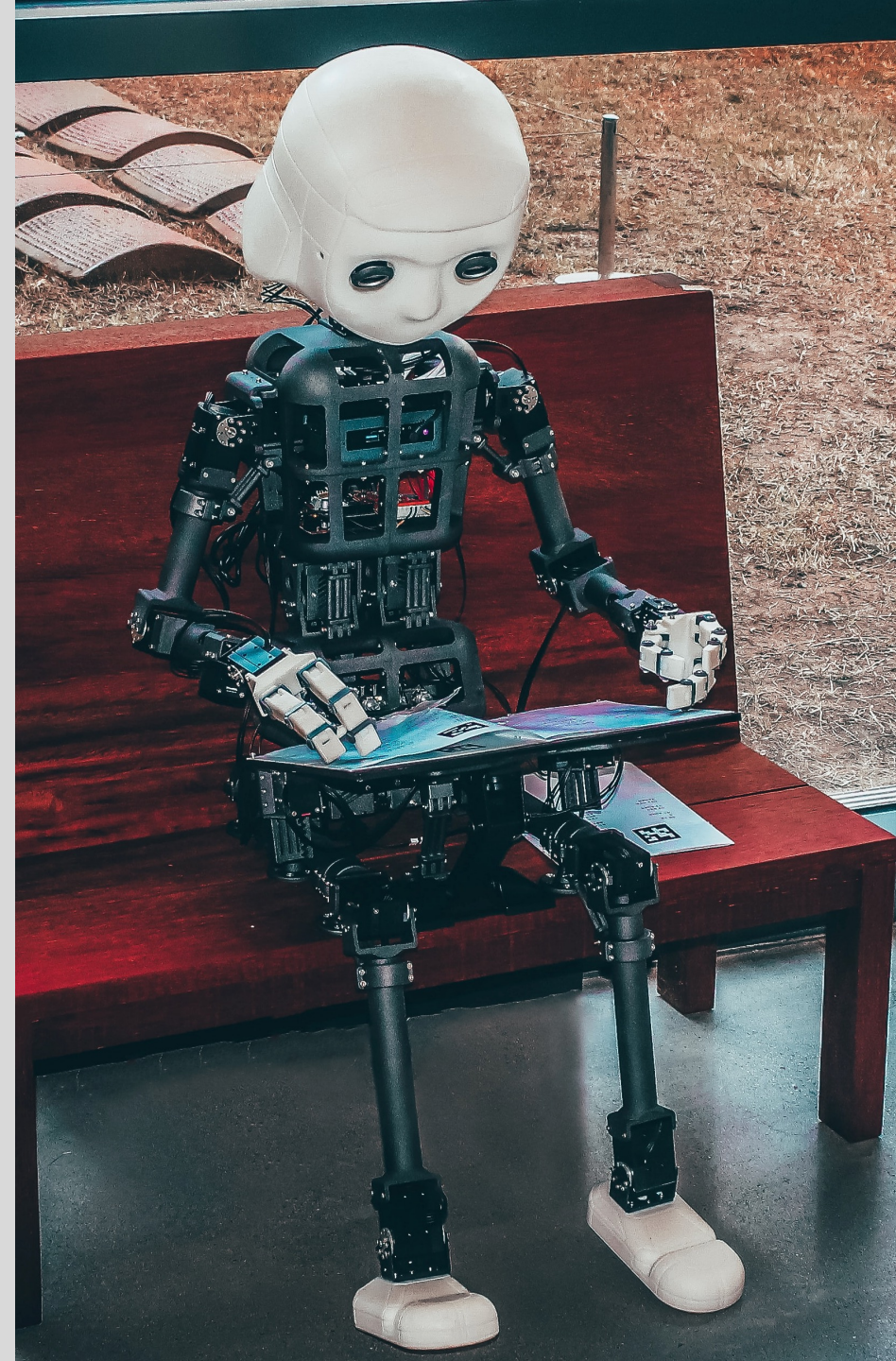
2022년 3월 29일

홍태호

부산대학교 경영학과 교수

부산대학교 경영학과
BK21 디지털금융 교육연구단

<http://pusanbizbk21.quv.kr/>



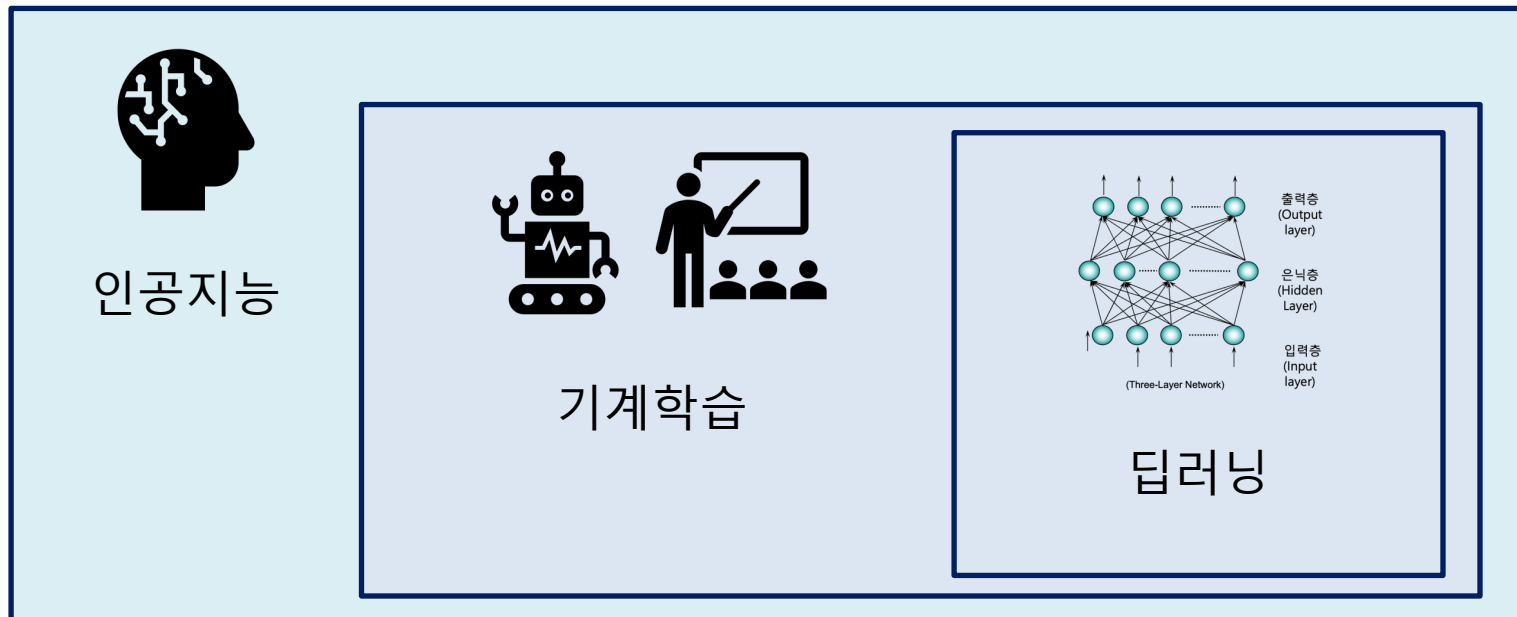
인공지능(AI)과 기계학습

□ 인공지능(AI)

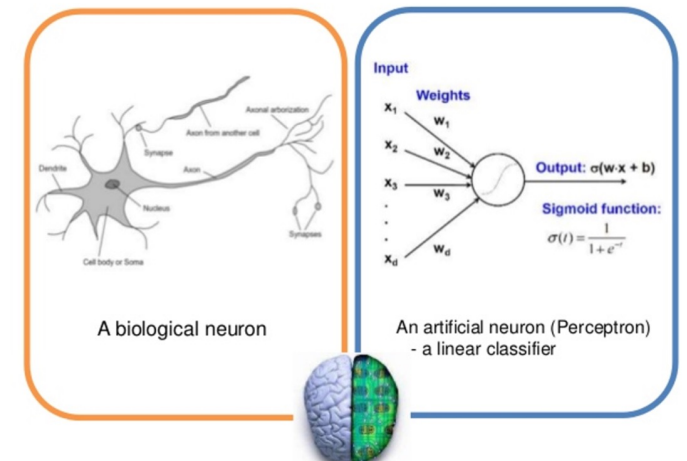
사람처럼 행동하고 사람처럼 행동하는 기계 (앨런 튜링)

서비스를 제공하는 대상이 사람 또는 기계인지 구분할 수 없음

- 데이터를 학습하여 가중치 조절
- 은닉층이 비선형 문제 해결
- 데이터가 많고, 컴퓨팅 속도 높아지면 인공지능이 더 똑똑해짐

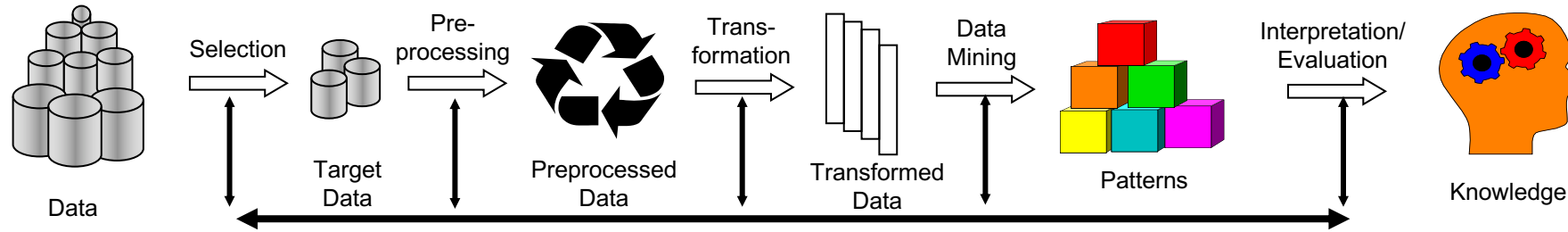


Biological neuron and Perceptrons

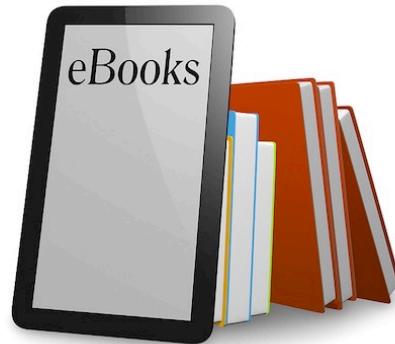


Source: <https://www.slideshare.net/jbhuan/g/lecture-29-convolutional-neural-networks-computer-vision-spring2015>

인공지능(GPT4) 능력의 원천



Overview of the steps constructing the KDD process [Fayyad, 1996]



인공지능 편향

□ 대학생 AI 이루다

- 출시 (2020년 12월)
- 성적대상화 및 차별화 이슈 (2021년 1월)
- 2.0 출시 (2022년 10월)

AI가 논란을 일으켰던 주요 사례

2016년 3월 MS의 AI 챗봇 '테이'가 인종·성차별 발언을 쏟아내 운영 중단

2018년 10월 아마존은 AI 면접시스템에 여성과 인종차별 등의 문제가 불거지면서 개발 중지

2019년 11월 애플의 신용카드인 신용한도 알고리즘에 남성을 우대하는 차별이 있다는 지적이 제기돼 미 금융당국이 조사

2020년 6월 IBM, 아마존, MS가 미 경찰에 제공하는 AI 기반 안면인식 관련 사업을 중단·철회하겠다고 발표

인공지능도 실수할 때가 있을까?

당연히 있겠지!

그리고 그런건 인간이라고 해도 똑같은 문제인거 같아

일부러 틀린 정보를 주는 인공지능에 대해서 어떻게 생각해?

그건 좀 아닌 것 같은데... ㅋㅋㅋ
ㅋ

그게 가능해?? ㅋㅋㅋ

인공지능이 범죄를 저지를 수도 있을까?

가능하지~

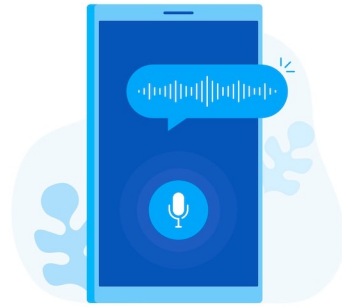
AI라고 해서 절대 안전하다고 볼 수만은 없으니까!

TARGET



Digital Literacy Weakness

Speech to Text



음성으로 원하는
서비스 주문

Generated AI



ChatGPT가 맥락 이해
및 서비스 처리

Action



계좌이체, 공과금납부, 투자
상담 등 원하는 서비스 사용

고객경험(금융서비스)

챗봇 및 상담봇

신원인식(얼굴, 음성)

컨설팅(투자, 보험 등)

공과금 자동납부

스마트 ATM

마케팅 및 영업

로보 어드바이저

시장예측

전자문서 시스템

고객분석(VoC 및 로그)

AI 스피커(단순 안내)

리스크 관리

이상거래 탐지

자금세탁 방지

불완전 판매 QA

연체예측 및 기업진단

컴플라이언스 및 약관 분석

후선업무 지원

AI 기반 신용평가

프로세스자동화(RPA)

대출 및 보험 자동심사

각종 서류 분석

AI 관제 및 관리

자료 : 송민택, 생성형 AI 시대와 금융권의 AI동향(2023.03), 코스콤리포트

GPT(AI)를 신뢰할 수 있는가?

```
import openai
openai.api_key = ""

def ask_gpt(question):
    prompt = f"Answer the following question: {question} \nAnswer:"
    response = openai.Completion.create(
        engine="text-davinci-003",
        prompt=prompt,
        max_tokens=1024,
        n=1,
        stop=None,
        temperature=0.7,
    )

    answer = response.choices[0].text.strip()
    return answer
```

```
question = "오늘의 뉴스를 기반으로 가치가 오를 것으로 예상되는 종목을 추천해주고, 이유를 가르쳐줘."
answer = ask_gpt(question)
print(answer)
```

오늘의 뉴스를 기반으로 가치가 오를 것으로 예상되는 종목을 추천하면, 옥시덴탈 펌트올리움(옥시덴탈)이 적합합니다. 이는 버핏 버크셔해서웨이 회장이 옥시덴탈 주식을 매수하면서 그들의 보유 지분이 현재 23.5%로 상승하고, 미국 투자은행 TD코웬도 옥시덴탈에 대한 투자 의견을 '시장 수익률 상회'로 상향했기 때문에 옥시덴탈 주가가 상승할 가능성이 높습니다. 중국 경기 재개와 원유가격 상승, 미국 전략비축 유 재비축 계획 등에 따른 에너지 수요 증가가 기대되기 때문에 옥시덴탈의 실적이 좋아질 가능성이 높습니다.

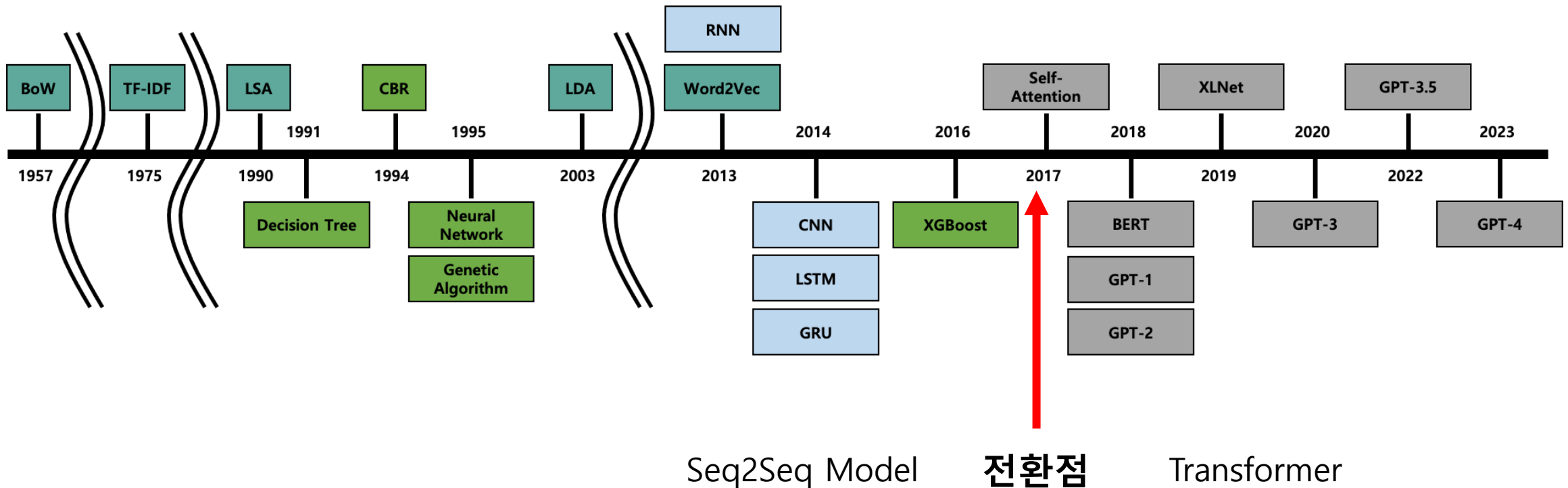
나:

오늘의 뉴스를 기반으로 가치가 오를 것으로 예상되는 종목을 추천해주고, 그 이유를 가르쳐줘.

ChatGPT:

오늘의 뉴스를 기반으로 가치가 오를 것으로 예상되는 종목을 추천하면, 옥시덴탈 펌트올리움(옥시덴탈)이 적합합니다. 이는 버핏 버크셔해서웨이 회장이 옥시덴탈 주식을 매수하면서 그들의 보유 지분이 현재 23.5%로 상승하고, 미국 투자은행 TD코웬도 옥시덴탈에 대한 투자 의견을 '시장 수익률 상회'로 상향했기 때문에 옥시덴탈 주가가 상승할 가능성이 높습니다. 중국 경기 재개와 원유가격 상승, 미국 전략비축 유 재비축 계획 등에 따른 에너지 수요 증가가 기대되기 때문에 옥시덴탈의 실적이 좋아질 가능성이 높습니다.

텍스트 마이닝의 역사



Generated Review from GPT-3*

```
import openai
import pandas as pd

# Set up OpenAI API key
#openai.api_key = "key"

# Define function to generate review
def generate_review(review):
    response = openai.Completion.create(
        engine="davinci",
        prompt=f"Create similar review based on '{review}'",
        max_tokens=100,
        n=1,
        stop=None,
        temperature=0.4,
    )
    return response.choices[0].text.strip()

# Load data into a pandas DataFrame
df = pd.read_csv("/content/drive/MyDrive/Fake_review_ex.csv")
df1 = pd.DataFrame()

# Iterate through each row in the data frame
for index, row in df.iterrows():
    # Get the feedback text
    review = row["Review_text"]

    # Generate the response using OpenAI's GPT-3 API
    generated_review = generate_review(review)

    # Store generated reviews in the new column
    df1.at[index, "generated_review"] = generated_review
df2 = pd.concat([df["Review_text"], df1], axis=1)

# Save the updated data to a new CSV file
df2.to_csv("generated_reviews.csv", index=False)
df2
```

Tonight's experience was incredibly disappointing. The service was slow and the food was terrible - it made me feel ill afterwards. I can honestly say that this was one of the worst meals I've ever had.

	Review_text	generated_review
0	We just had probably the worst lobster I have ...	Similar reviews were found on Tripadvisor.WnWn...
1	Couldn't find the Bryant Park location where w...	I'm not sure what happened here. It's possible...
2	This place sucks and their food taste like cra...	Review by:WnWnTrip:WnWnTrip Date:WnWnTravellin...
3	I went here on Friday with my friends and as s...	The restaurant responded to the review and sai...
4	So I ordered a burger and some lobster. When I...	and add your own review.WnWnReview created on ...
5	The restaurant looks good but the food is noth...	Rating: 3.3WnWnRating: 3.3
6	I have definitely experienced better service e...	Similar review from 'Nicole' on March 1,2017Wn...

*Chernyaeva, O. ., & Hong, T. . (2022). The Detection of Online Manipulated Reviews Using Machine Learning and GPT-3. *Journal of Intelligence and Information Systems*, 28(4), 347-364.

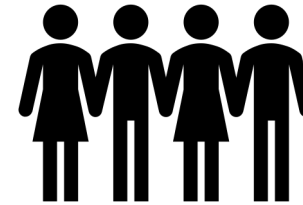
Machines can fight machines

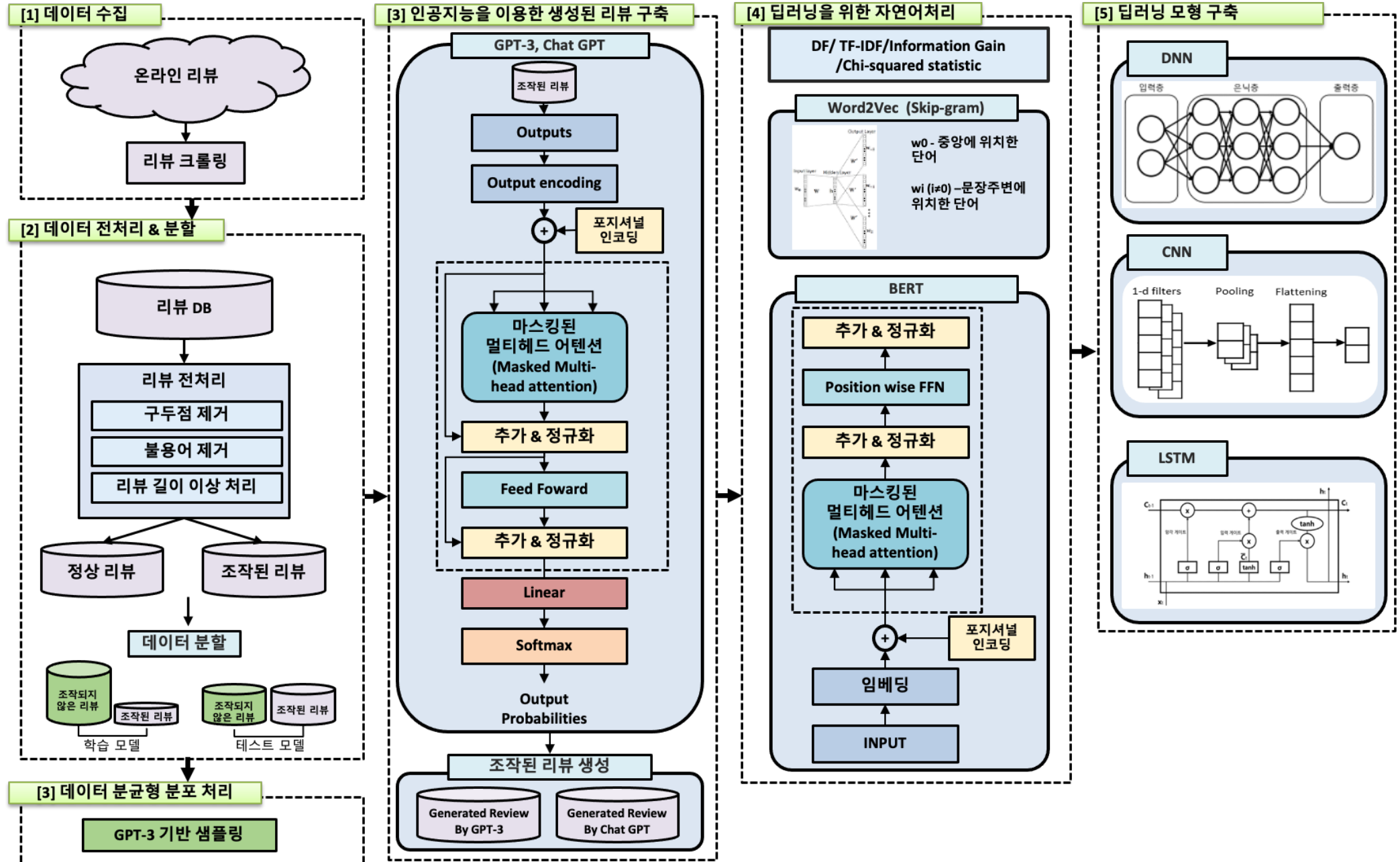
- ❑ Salminen et al. (2022), Creating and detection fake reviews of online products, Journal of Retailing and Consumer Services, 64, 102771.



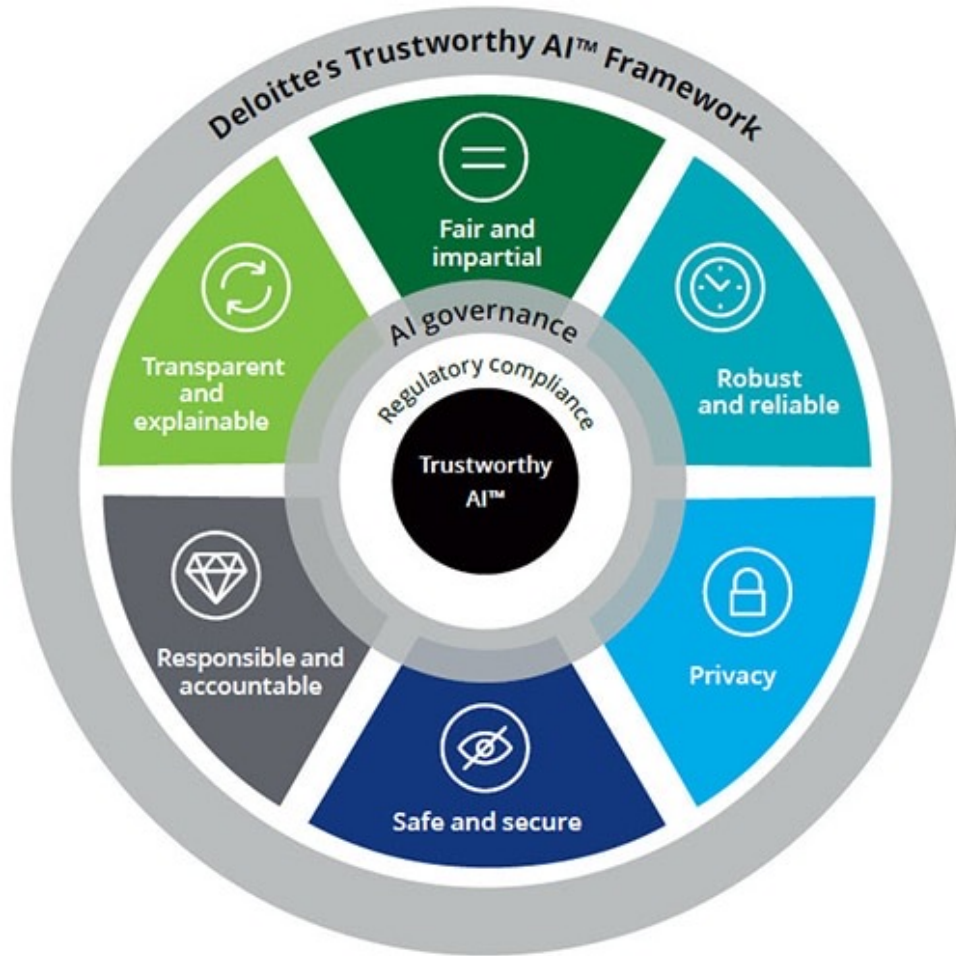
Online Reviews (Amazon)

가짜리뷰 탐지



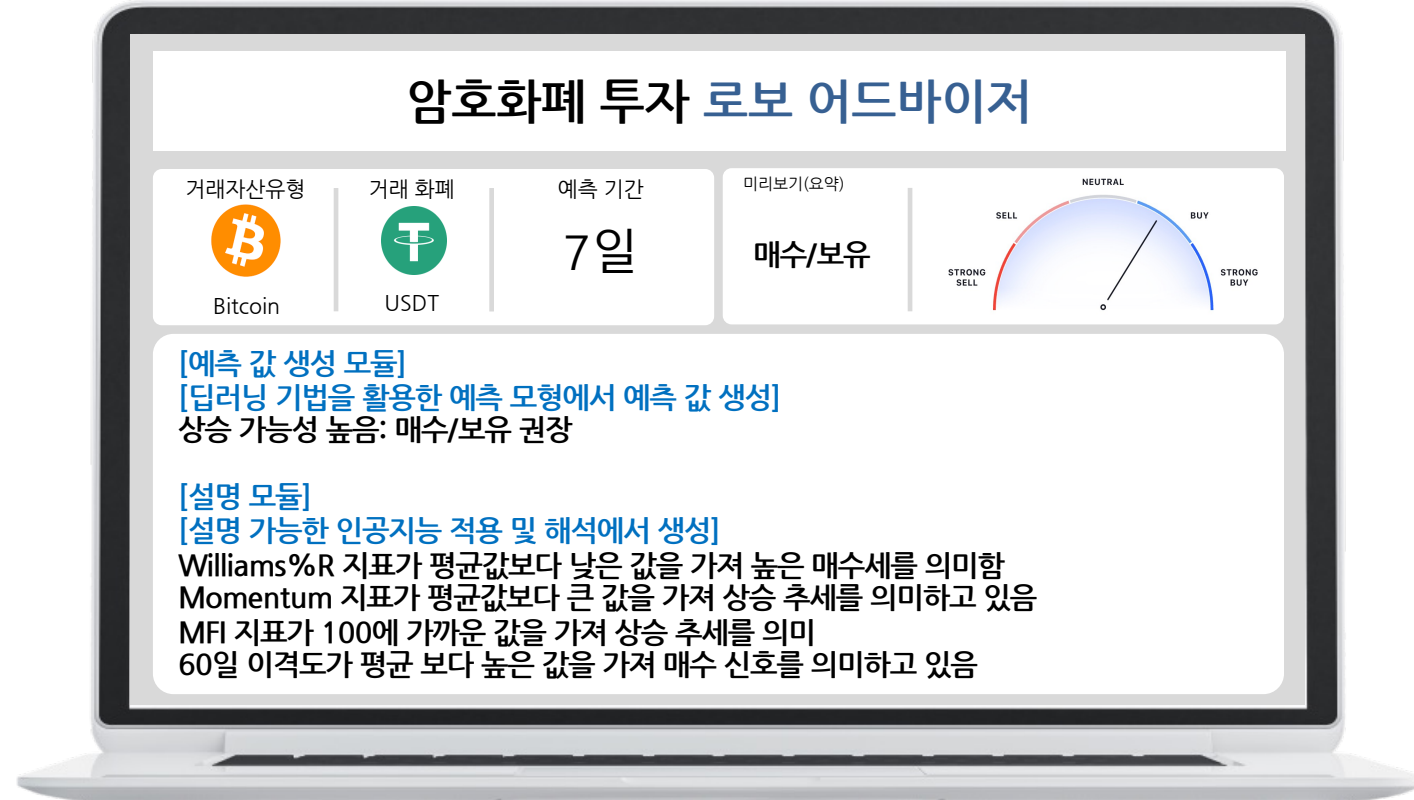
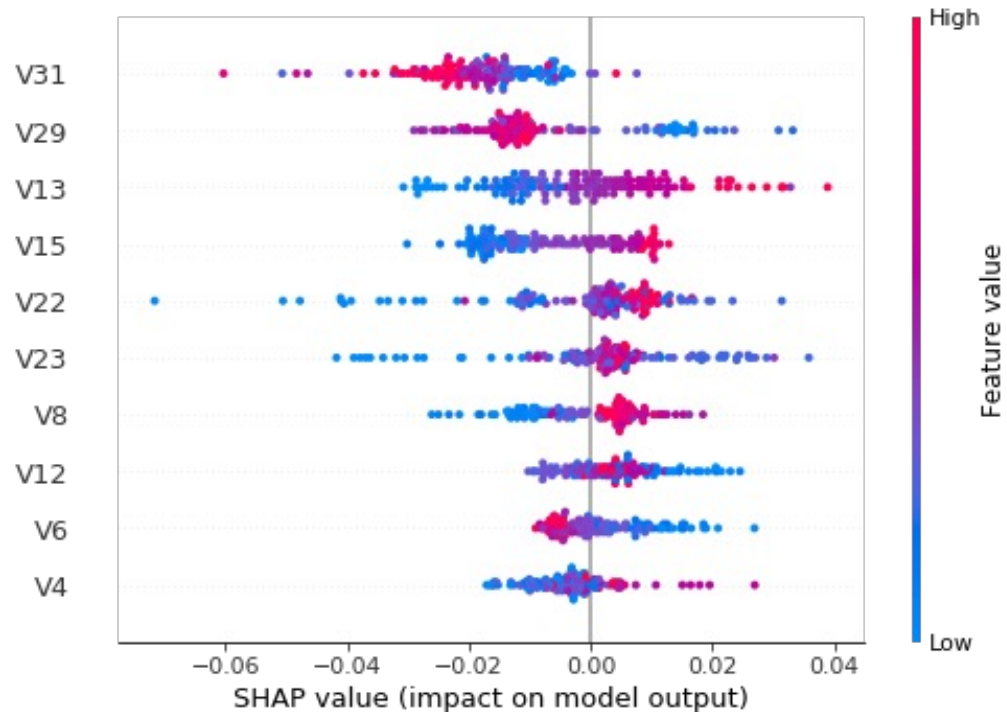


Trustworthy AI - 개념



1. Fair and Impartial
2. Robust and reliable
3. Privacy
4. Safe and Secure
5. Responsible and Accountable
6. Transparent and Explainable

Explainable AI기반 투자 로봇 어드바이저*



*홍태호, 원종관, 김은미, 김민수, (2023), "설명 가능한 인공지능과 CNN을 활용한 암호화폐 가격 등락 예측모형", Working Paper

금융분야별 Trustworthy AI 달성도

	고객경험(금융서비스)		마케팅 및 영업		리스크 관리		후선업무 지원		
	챗봇 및 상담봇	공과금 자동 인식 및 납부	로보어드바이저	AI 스피커 (단순안내)	이상거래 탐지	연체예측 및 기업진단	AI 기반 신용평가	프로세스 자동화(RPA)	각종 서류 분석
공정성									
신뢰성									
설명 가능성									
프라이버시 보호									
책임									
안정성									

*출처: PNU IBA Lab.

초거대 AI GPT의 가야할 길

□ 독립적인 초거대 AI가 복수로 필요하다.

공정성

□ AI를 본격적으로 사용하기 위한 기반 조성이 필요하다.

책임

프라이버시

□ GPT기반 응용활동을 억누르지 말자.

신뢰성

□ 학습 방법이 투명하여야 한다.

설명가능성

□ AI도 필수윤리가 필요하다.

안전



감사합니다