

AI 이노베이션

데이터 전략

정두희 교수
한동대학교 ICT창업학부

중간고사 공부방법

안내...

AI 이노베이션 아카이브

AI 혁신 준비

- AI 기술 개괄
- AI 혁신의 속성
- 문제 정의

AI 기능 혁신

- AI 5대 기능
- AI기능 적용방법

AI 기술 구현

- AI 알고리즘
- AI 데이터

AI 가치 창출

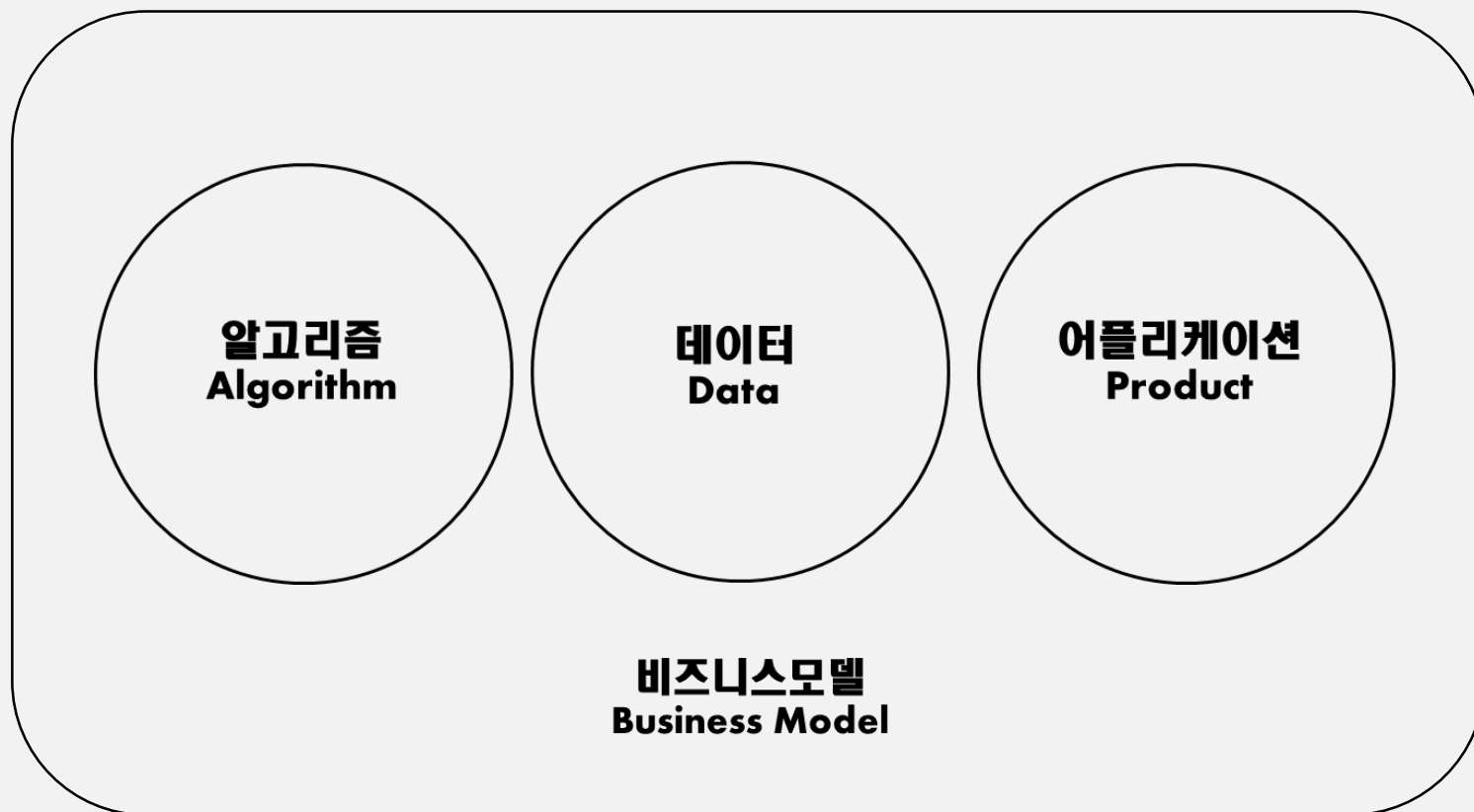
- AI 가치경로
- AI 학습루프

Warm-Up

“AI 시스템을 개발하는 기업들에게 흔히 나타나는 실수가 정교한 예측 알고리즘을 개발해놓고, 여기에 필요한 데이터를 가지고 있지 않는다는 것이다” - 마이크로소프트 CTO 제이콥 스폐엘스트라

선진기업의 API 공유

Warm-up

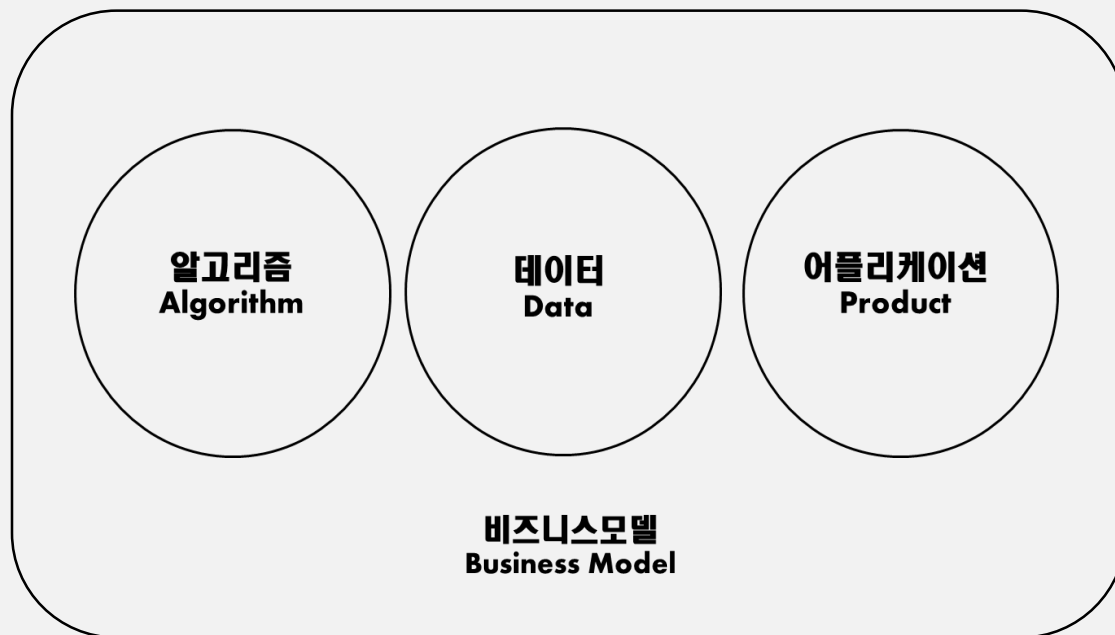


LO

- AI 혁신에서 데이터의 역할을 이해
- 데이터 경쟁력을 높이기 데이터 확보 전략을 이해
- 이를 바탕으로 AI 혁신을 디자인하기 위한 기술명세서를 개발

KEY: 핵심 데이터의 정의

- ✓ AI 혁신의 성과는 데이터 확보능력과도 밀접한 관련
- ✓ 데이터 확보를 위해서는 먼저 학습에 필요한 핵심 데이터는 무엇인지 정의



KEY: 데이터 소스는 어디에?

- ✓ 정의내린 데이터를 어디에서 확보할 수 있는지 확인
- ✓ 전방위적인 데이터 확보 필요

“넷플릭스에서는
데이터를 어떻게
수집하고 있을까?”

데이터 소스 예시

- * 회사 내부의 데이터
- * API 통한 외부 데이터
- * 웹 스크랩 통한 외부 데이터
- * 제휴기업 데이터베이스
- * 공공데이터

KEY: 좋은 데이터의 조건

✓ 좋은 데이터와 안좋은 데이터 구분 필요

✓ 좋은 데이터의 조건은?

- * 데이터 크기
- * 데이터 시의성 및 생성의 지속성
- * 데이터 다양성
- * 데이터 편향성
- * 적절한 속성의 정의(레이블링)
- * 데이터 관리 시스템

KEY: 좋은 데이터의 조건

✓ 좋은 데이터와 안좋은 데이터 구분 필요

✓ 좋은 데이터의 조건은?

- * 데이터 크기
- * 데이터 시의성 및 생성의 지속성
- * 데이터 다양성
- * 데이터 편향성 無
- * 적절한 속성의 정의(레이블링)
- * 데이터 관리 시스템

KEY: 데이터 부재 이슈

Poll

Q. 데이터가 존재하지 않을 경우
어떻게 학습을 시킬까?



KEY: 데이터 부재에 대한 해결책

데이터적 접근: 학습데이터 생성플랫폼

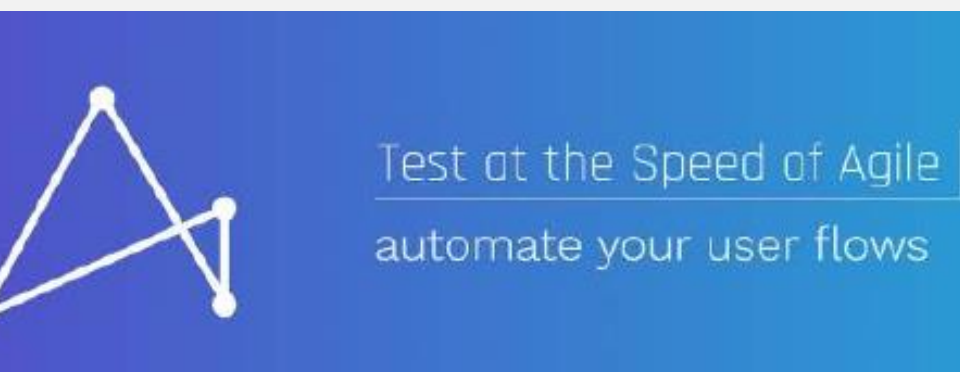
API 통해 학습용 Raw데이터 확보, 실제 학습에 사용할 기준 데이터로 반환
정확도 95% 이상의 기준 데이터를 받을수 있으며, 개발 속도 증대
학습용 데이터의 최초 수집 이후 '정제'에 초점



KEY: 데이터 부재에 대한 해결책

데이터적 접근: 테스트봇

- 인공지능 봇(테스팅봇)이 사람처럼 서비스의 각 요소를 인식하고 실행
- 사용자 경험 및 퍼포먼스 데이터 생성
- 테스트 봇을 통한 결과와 인간 테스트 결과 간의 유사도가 중요



KEY: 데이터 부재에 대한 해결책

알고리즘적 접근: 강화학습

Reinforcement learning

- 데이터가 주어지지 않으며 알고리즘이 스스로 다양한 시도를 하면서 최적 결과 도출
- 핵심적인 규칙만 익히고 바로 실전에 들어가고, 실전을 통해 학습
- 시행 착오(Trial-and-error)와 지연 보상(delayed reward) 등 보상을 통한 성과 유도



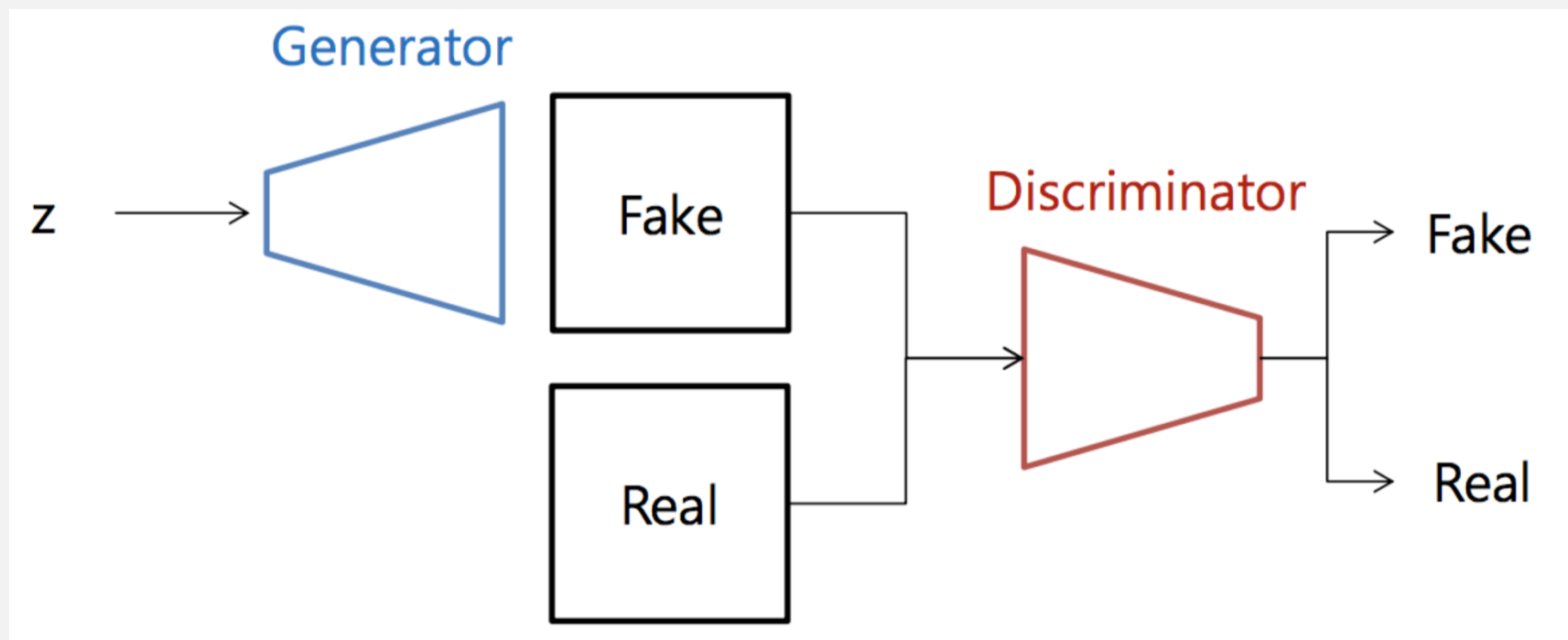
KEY: 데이터 부재에 대한 해결책

알고리즘적 접근: 생성적 적대 신경망

Generative Adversarial Network, GAN

인간 개입이나 추가적 학습 데이터 없이 AI 스스로 학습하게 하는 신경망 기법,
서로 대립되는 두 기능이 상호경쟁을 통해 성능을 높이는 비지도학습법

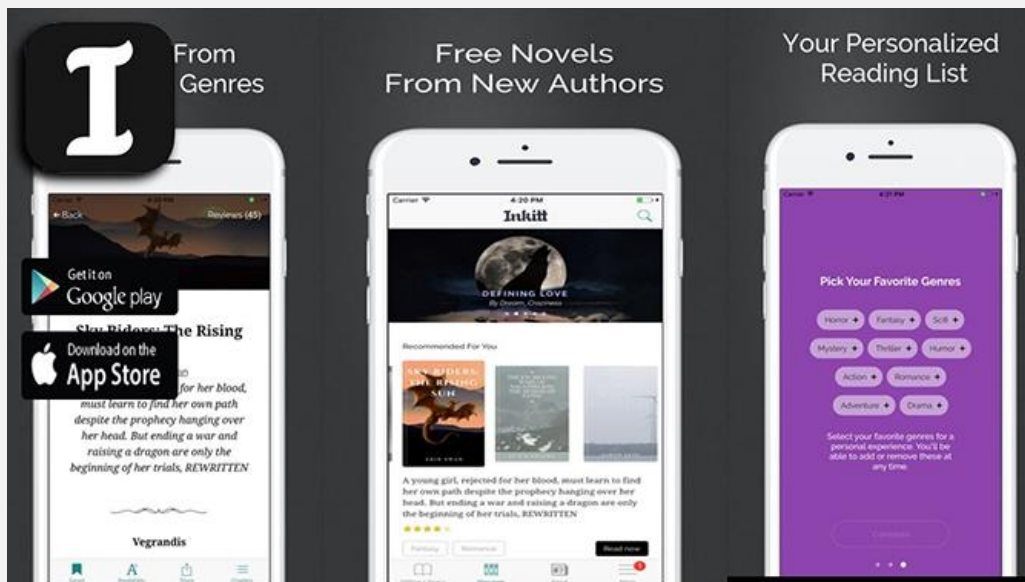
Ex. 위조지폐 만드는 기능 vs. 지폐 위조 여부 감지 기능 → 두 기능의 경쟁으로 성능 향상



KEY: 데이터 부재에 대한 해결책

비즈니스모델적 접근

- AI 출판사 인키트: 사전연재 방식으로 시장학습 및 흥행 예측
- 루트 인슈어런스: 2~3주 정도 시험운전 통해 운전자의 운전습관과 사고 위험성을 예측, 보험료 산정



ROOT

ACTIVITY: AI 혁신 기술명세서

Aii

알고리즘을 선택하기
위한 질문

1

테크닉 검토

AI 기능을 수행하기 위한
세부 테크닉은 무엇인가?

2

작업속성 검토

데이터크기, 긴급성, 정확성 등
작업 속성은 무엇인가?

3

알고리즘 선정

이 속성에 따른 최적의
알고리즘은 무엇인가?

ACTIVITY: AI 혁신 기술명세서

AI 혁신 기술명세서를 작성 (ex. 이메일 스팸메일 분류 어플 개발)

문제	핵심기능	테크닉	작업속성	알고리즘
이메일에서 스팸메일 자동 분류 필요	·스팸인식(맥락인식) ·스팸분류(자동처리)	·텍스트인식 ·분류 (Classification)	· 메 일 로 이 름 인 데이터가 방대 · 스 팸 메 일 분류작업은 수신과 동시에 실시간 처리 필요	나 이브 베 이즈 분류(Naïve Bayes) 알고리즘
				데이터
				·정상 메일과 스팸 메일이 섞인 메일 데이터로 학습 · 데이터 플랫폼 (ex.캐글)에서 제공하는 메일 데이터 확보

LO

- 반드시 알아야 할 주요 알고리즘 이해
- 알고리즘 직접 체험

중간고사 안내

중간고사 출제 범위

1. 인공지능 특징점과 한계점
2. AI 5대 기능
3. 머신러닝 체계
4. 딥러닝
5. 자연어처리
6. 생성적 적대 신경망
7. 데이터 확보

10주차 금요일 수업시간 온라인 접속

다음주 공지

- * 프로포절 발표 11주차(미평가)
- * 다음주 동영상 공지
AI 이노베이션 전체 복습용 영상