Data Science

Data Ethics & Trends

노기섭 교수

(kafa46@hongik.ac.kr)

Lecture Goals

- 데이터 윤리가 왜 중요한지 이해
- 개인정보 보호 규제(GDPR, CCPA, 한국 개인정보보호법) 비교
- 알고리즘 편향의 원인·사례·해결 전략 이해
- 데이터 거버넌스
- AutoML, MLOps, 생성형 AI(Generative AI) 최신 트렌드 파악
- 실제 데이터/AI 시스템에서 윤리가 어떻게 고려되는지 정리

개 요

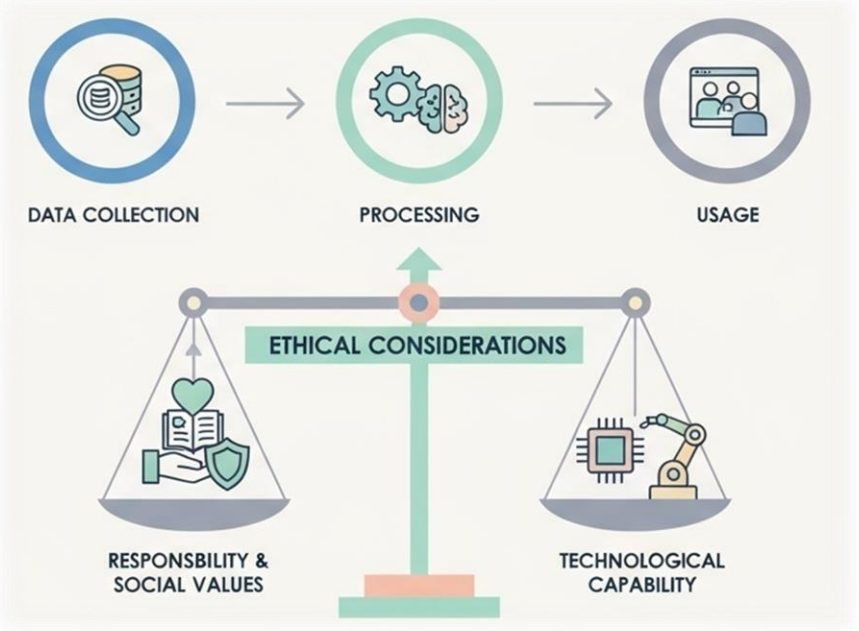
Why Data Ethics?

데이터 윤리란?

- 데이터를 수집·처리·분석·활용하는 전 과정에서 "사회적으로 책임 있는 방식"을 지키는 것

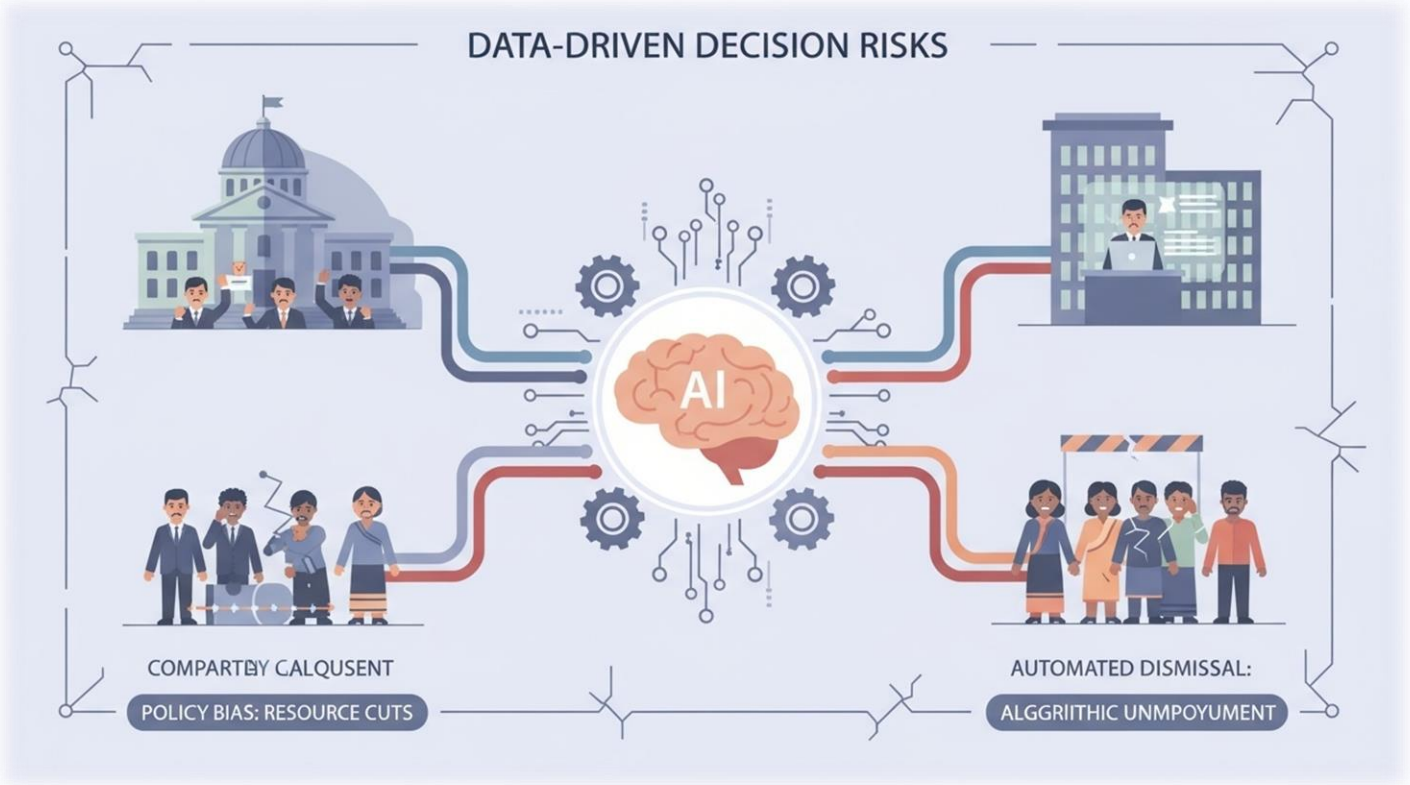
왜 중요한가?

- 기업·정부의 자동화된 판단이 사람의 삶에 큰 영향을 미침
- 데이터 기반 의사결정이 공정성과 신뢰성을 결정
- AI가 의사결정에 직접 참여하는 시대



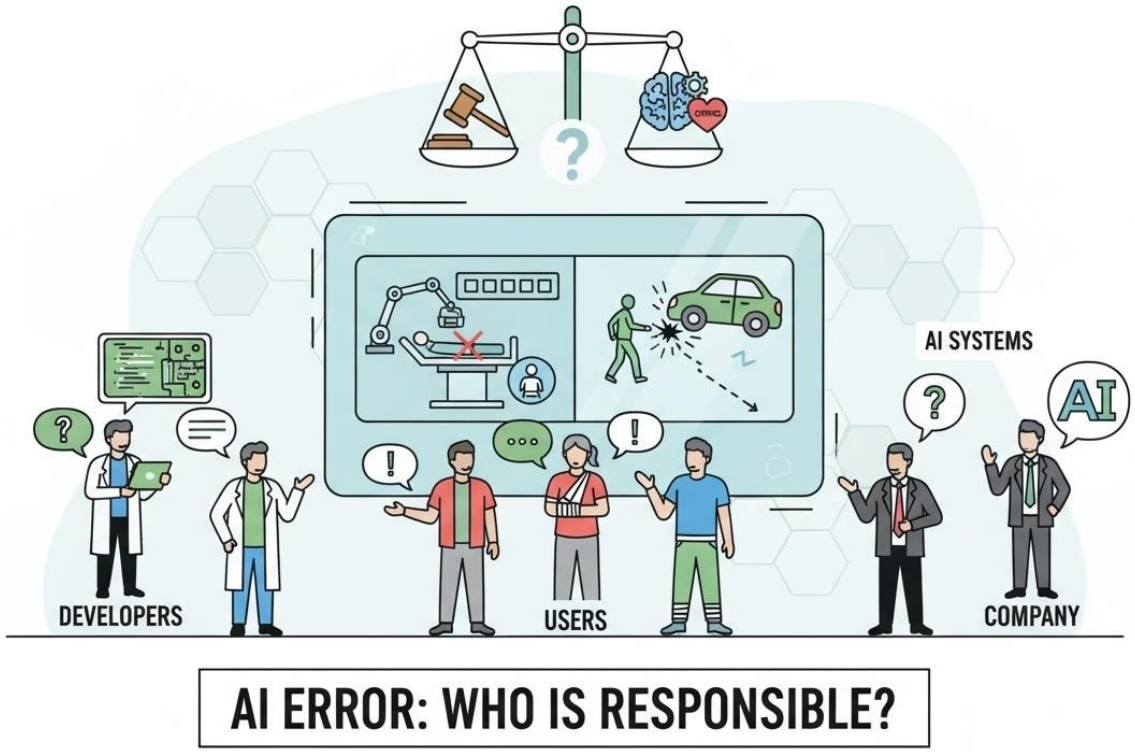
데이터 기반 의사결정이 사회에 미치는 영향 확대

- 공공 정책 결정 시 특정 지역·계층에 불리한 데이터 기반 판단이 내려질 위험
- 기업의 자동화된 의사결정 시스템이 인간적 판단 없이 해고·대출 거절 등을 결정하는 사례 증가



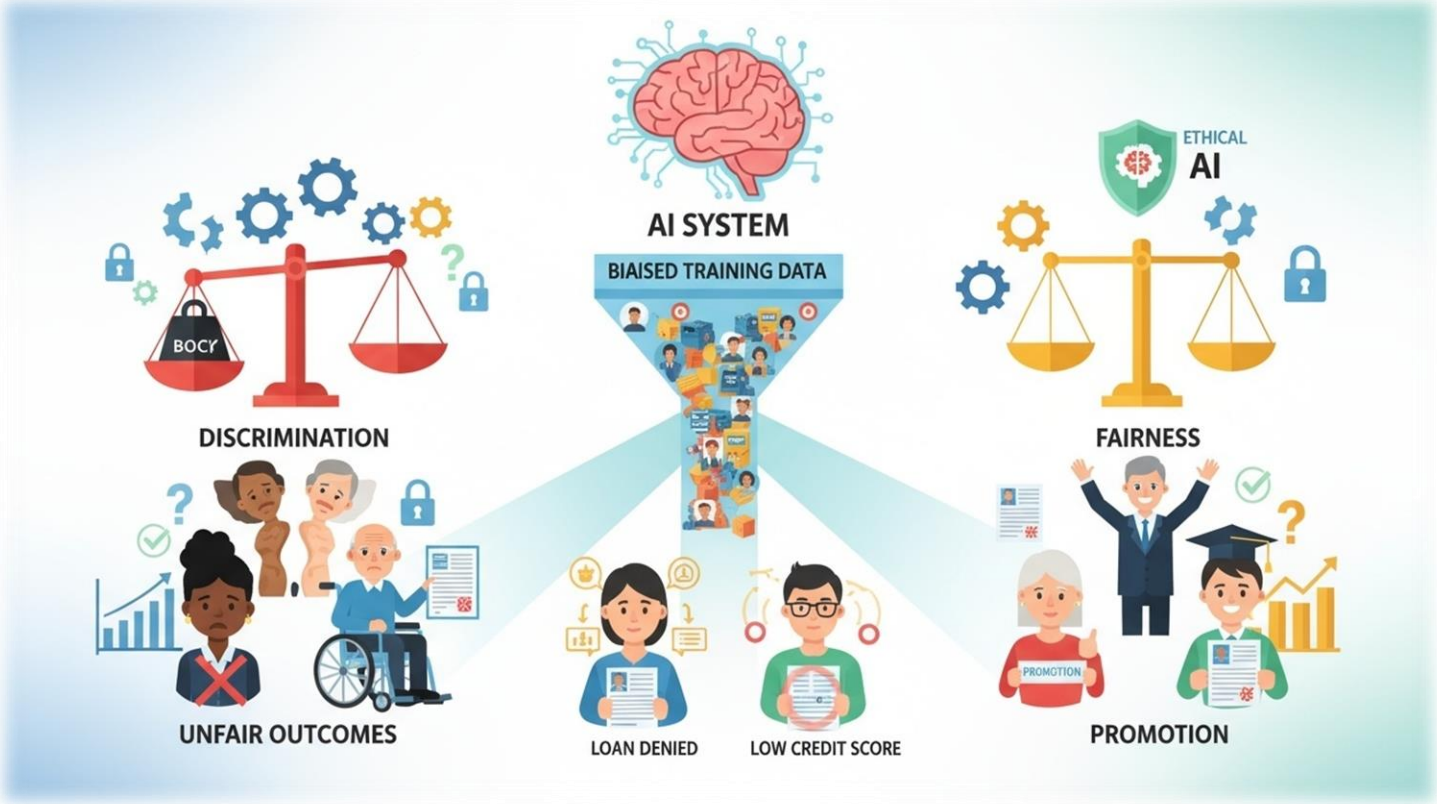
AI 모델의 결정에 대한 책임 소재 문제

- AI가 잘못된 판단을 내렸을 때 책임이 개발자, 사용자, 기업 중 누구에게 있는지 불분명
- 의료·자율주행 등 생명에 영향을 미치는 분야에서 AI 오류 발생 시 법적 분쟁 심화



개인정보 침해 및 차별적 알고리즘에 대한 우려 증가

- 동의 없이 수집된 민감 정보가 제3자에게 판매되거나 마케팅에 악용
- 학습 데이터 편향으로 인해 인종·성별·장애 여부 등에 따른 차별적 결과가 지속적으로 재생산



Data Privacy

개인정보란?

개인정보란?

- 개인을 식별할 수 있는 모든 정보
  - 예: 이름, 연락처, 위치정보, 쿠키/로그, 얼굴·음성 데이터 등

주요 이슈

이슈	설명	사례
무단 수집	동의 없이 데이터 수집	위치추적 앱, 웹 쿠키
목적 외 사용	동의 범위 초과 활용	마케팅 재활용
데이터 유출	보안 관리 미흡	금융/의료 정보 유출

주요 법/정책 비교표

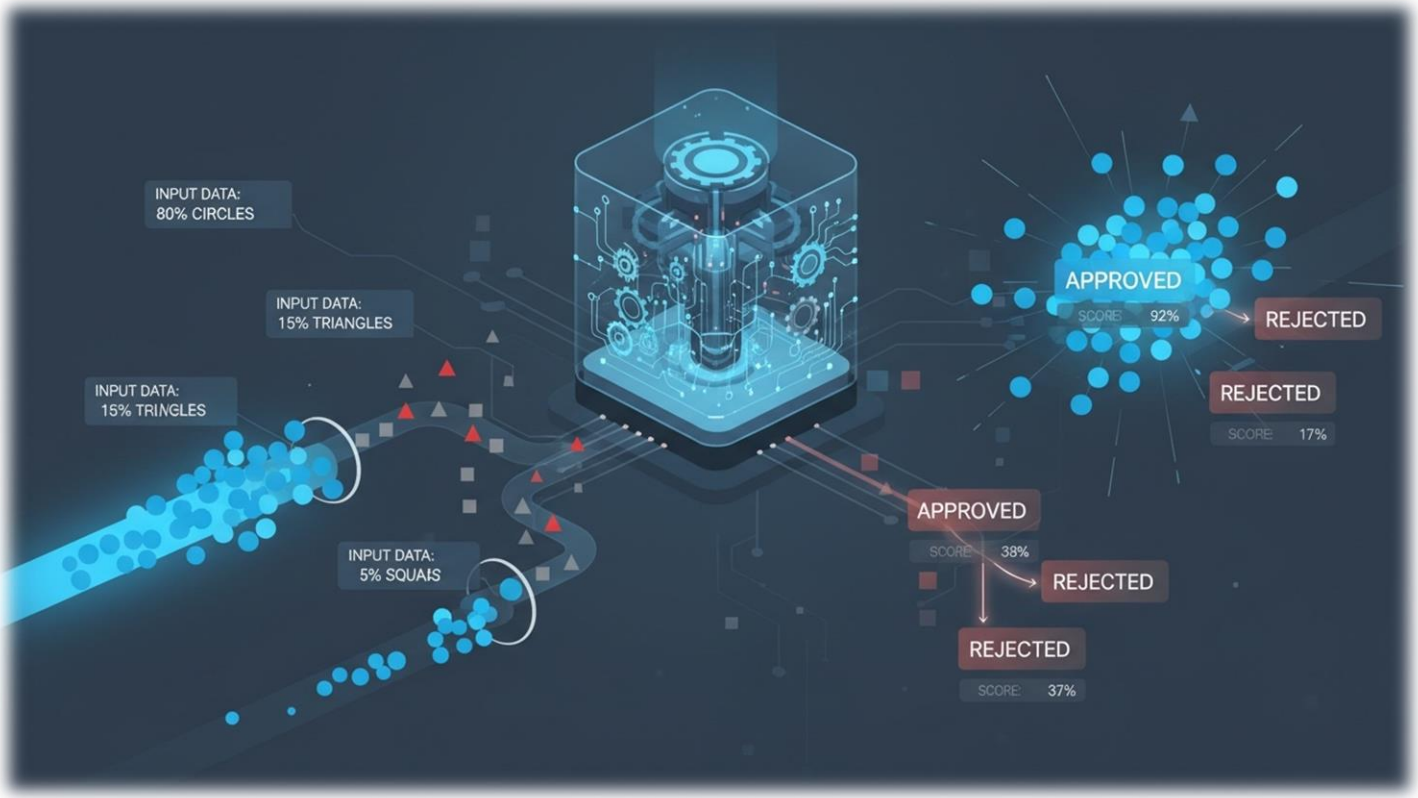
법/정책	지역	특징	주요 내용	홈페이지
GDPR	유럽 (EU)	가장 강력한 개인정보 보호 규제, 글로벌 표준 역할	• 데이터 처리 최소화 원칙 • 개인정보 처리에 대한 명확한 동의 필요 • 잊혀질 권리/데이터 이동권 보장 • 위반 시 매출의 최대 4% 또는 2천만 유로 과징금	gdpr.eu
CCPA	미국 (캘리포니아)	소비자에게 데이터 사용 통제권 강화	• 수집 정보 공개 요구권 • 개인 데이터 판매 거부권 (Do Not Sell My Info) • 개인정보 삭제 요청권	oag.ca.gov/privacy/ccpa
개인정보 보호법	대한민국	개인정보 보호와 정보 활용 균형 추구	• 수집·보관·이용·파기 절차 명확화 • 민감정보 및 고유식별정보 보호 강화 • 개인정보 영향평가 수행 • 위반 시 행정처분 및 과징금 부과	pipc.go.kr

알고리즘 편향

알고리즘 편향 (Algorithmic Bias)이란 ?

알고리즘 편향 (Algorithmic Bias)이란 ?

- 데이터나 모델 설계의 불균형으로 인해 특정 집단에 불공정한 결과가 발생하는 현상



알고리즘 편향 원인

원인

- 편향된 학습 데이터
- 평가 지표나 목표 설정 시 사회적 맥락 미반영
- 설계자/개발자의 주관이 모델에 반영

사례

분야	편향 사례
채용	특정 성별/학교 선호
금융	낮은 소득 지역에 일괄 신용점수 하향
의료	특정 인종 데이터 부족으로 진단 정확도 편차 발생
얼굴 인식	백인 남성 대비 유색인종 여성 오류율 ↑

대응 전략 1.

■ 데이터 다양성 확보

- 대표성(Representativeness) 점검
  - 표본이 전체 모집단(성별·연령·지역·장애 여부 등)을 얼마나 반영하는지 분포 비교
- 수집·보강(Collect & Enrich)
  - 능동 학습(Active Learning): 불확실성이 큰 샘플 위주로 추가 라벨링
  - 데이터 보강: 이미지(밝기/각도/배경제거), 텍스트(역번역), 음성(속도/피치) 등 다양하게 증강
- 재가중/재샘플링(Reweight/Resample)
  - 그룹별 가중치 부여로 학습 손실에 공정성 반영
- 라벨 품질 관리
  - 주석자 간 일치도 측정
  - 모호 케이스에 대한 라벨 정책 문서화(라벨 가이드라인, 예외 규칙)
- 데이터 문서화·거버넌스
  - Datasheets for Datasets, Model Cards 작성
  - 수집 목적, 법적 근거, 보관·파기 정책, 재사용 범위 명시

대응 전략 2. 모델 Explainability 적용 (SHAP, LIME 등)

목표

- 예측 근거를 투명하게 제시해 편향 원인(특정 속성·특징 과대 의존)을 조기에 탐지
- 이해관계자(사용자, 심사자, 규제기관) 커뮤니케이션에 활용

왜 설명 가능성이 필요한가?

- AI 모델이 어떤 이유로 그런 예측을 했는지 설명할 수 있어야,
- "어디서 편향이 생겼는지, 무엇이 문제인지"를 찾고 개선할 수 있음

실무 적용 팁

- 금지/민감 속성(Protected Attributes)이 직접·간접적으로 중요한 피처로 드러나면 제거·변환·규제를 고려 (예: 우편번호가 인종을 대리하는 경우) 등

대응 전략 3. 편향 측정 지표 도입

평가 기본 원칙

- 정확도만 보지 말고, 집단별 성능 격차를 함께 본다.
- 집단 A, B에 대해 TPR/FPR/Precision/Recall/Calibration을 동일 조건에서 비교
- 운영 임계값(Threshold)을 집단별로 조정하는 Post-processing도 고려

주요 공정성 지표 개요

- **Demographic Parity (DP):** 긍정 예측 비율 자체가 집단 간 비슷해야 한다.
- **Equalized Odds (EO)**: TPR, FPR이 집단 간 유사해야 한다.
- **Equal Opportunity (EOp)**: 양성에 대한 기회 평등해야 한다.

대응 전략 4. 운영 관점 통합 (MLOps + 거버넌스)

파이프라인 내재화

- 데이터 검수(분포·라벨 품질) → 공정성 지표 계산 → Explainability 리포트 → 배포 게이트 구성
- CI/CD 파이프라인에 공정성 실패 시 배포 중단 규칙을 포함

모니터링

- 집단별 성능·에러·확률 보정(Calibration) 추적 대시보드 운영
- 실시간 경보와 롤백·재학습 트리거를 연동

대응 전략 4. 리스크 컴플라이언스 (1/2)

컴플라이언스(Compliance)

- 조직이 법률, 규정, 계약, 내부정책을 준수하도록 설계·운영·감사하는 활동 전반을 의미
- 개인정보보호법, GDPR, CCPA 같은 법규와 고객사 SLA·DPA, 내부 보안·개인정보 정책 준수가 포함
- 목표: 과징금·소송 등 법적 리스크와 평판 리스크 ↓, 데이터/AI 운영의 투명성 ↑

핵심 구성요소

- **정책**(Policy): 데이터 수집·이용·보관·파기, 제3자 제공, 재식별 금지 등을 문서화
- **프로세스**(Process): 승인, 변경관리(모델·데이터·코드), 사고 대응, 사용자 요청 처리 흐름 정의
- **증빙**(Artifacts): 로그, 보고서, 승인 기록, 교육 이력 등 준수 사실을 입증할 자료
- **감사**(Audit): 정기 점검, 내부 감사, ISO 27001·SOC 2 등 외부 인증 수행

대응 전략 4. 리스크 컴플라이언스 (2/2)

■ 감사 추적성(Auditability) 확보

- 데이터·모델 변경 이력, 승인자, 근거 문서를 기록해 누가·언제·무엇을·왜 변경했는지 재구성 가능해야 한다.

■ 사용자 설명·이의제기 프로세스

- 목적: 사용자가 결과를 이해하고 부당함을 제기할 수 있도록 투명한 설명과 수정 경로를 제공
- 프로세스 설계 예시
  - **접수 채널**: 웹 양식, 이메일, 고객센터
  - **SLA**: (예) 7영업일 내 1차 회신, 30일 내 최종 답변
  - **설명 제공**: 모델 카드(목적, 데이터 범주, 한계, 공정성 지표)와 사례 설명을 이해하기 쉽게 전달
  - **이의제기 처리**: 수동 재검토 후 필요 시 임계값 조정·재평가

    → 결과 안내, 반복 이슈는 정책·모델 개선 루프에 반영
  - **기록 관리**: 접수부터 개선 조치까지 티켓 시스템으로 추적

데이터 거버넌스

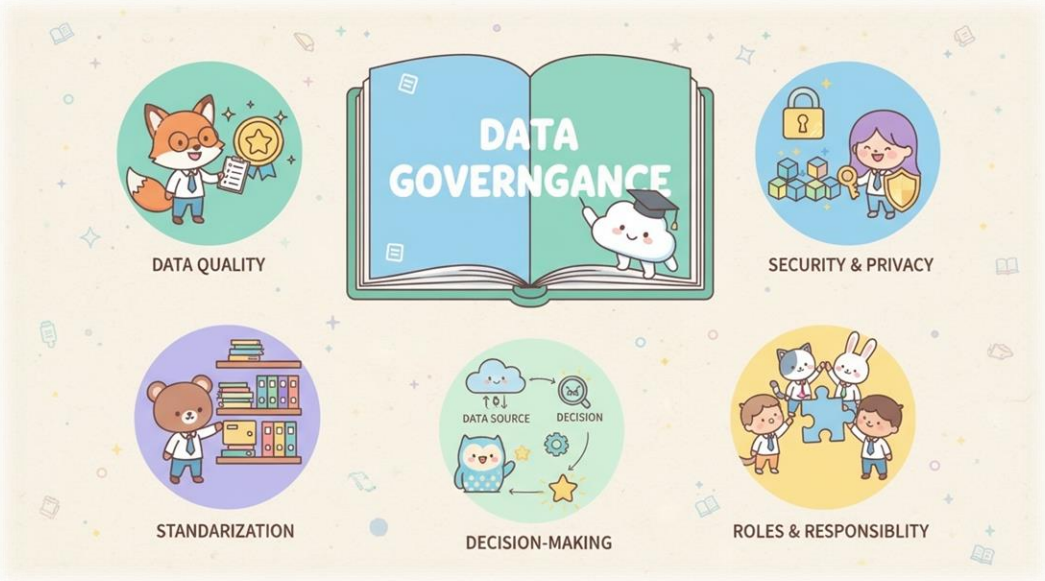
데이터 거버넌스

데이터 거버넌스(Data Governance)

- 조직이 보유한 데이터를 올바르고 안전하게 관리하고 활용하기 위한

  정책, 기준, 절차를 정하는 체계를 의미

데이터를 누가, 어떻게, 어떤 규칙에 따라 저장·사용·공유해야 하는지를 정해 데이터의 품질, 보안, 책임성을 보장하는 것!



데이터 거버넌스의 핵심 목표

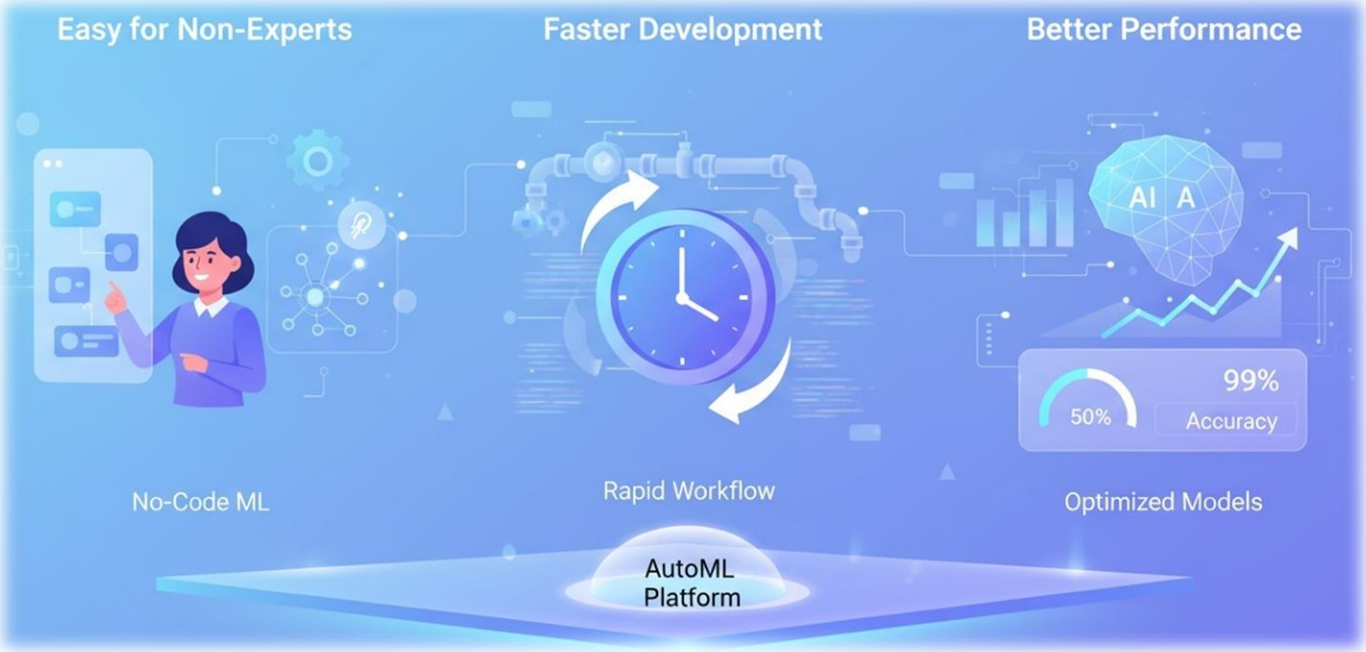
- **데이터 품질 확보**
  - 정확하고 신뢰할 수 있는 데이터를 유지
- **보안 및 개인정보 보호**
  - 데이터 접근 권한 관리, 유출 방지, 법적 준수
- **표준화**
  - 조직 전체가 동일한 데이터 정의와 규칙을 사용하도록 정착
- **책임과 역할 분배(R&R)**
  - 데이터와 관련된 업무에 대한 명확한 담당자 지정
- **데이터 활용 촉진**
  - 데이터가 효율적이고 가치 있게 활용될 수 있는 환경 조성

AutoML

AutoML(Automated Machine Learning)

AutoML(Automated Machine Learning)

- 머신러닝 모델 개발 과정에서 사람이 직접 수행하던 작업들을 자동화해주는 기술
- 데이터 전처리, Feature Engineering, 모델 선택, 하이퍼파라미터 튜닝, 성능 평가 등을 자동 수행
- 전문가가 아니더라도 효율적으로 고품질 ML 모델을 만들 수 있도록 돕는 시스템



AutoML 목적 및 장점

AutoML 핵심 목적

- 모델 개발 과정을 자동화하여 시간과 비용을 절약
- 복잡한 알고리즘 선택과 튜닝을 자동으로 수행해 모델 성능 향상
- 비전문가도 쉽게 모델을 만들 수 있도록 지원해 AI 활용 장벽을 낮춤

장점

- 머신러닝 비전문가도 쉽게 모델을 만들 수 있음
- 데이터 사이언티스트가 직접 수행하던 작업을 자동화함으로써, 보다 빠르게 모델을 실험, 결과확인
- 많은 후보 모델과 설정을 탐색해 최적화하기 때문에,

  사람의 직관에만 의존했을 때 놓칠 수 있는 높은 성능의 모델을 발견

대표 도구

도구	간단 특징	쉬운 설명 (한 줄 비유)
Google AutoML	GUI 기반 비전문가도 사용 쉬움 GCP 연동	"요리 초보도 버튼만 누르면 요리가 완성되는 자동 조리기 같은 AutoML"
Auto-Sklearn	오픈소스 Scikit-Learn 기반 자동 모델 탐색	"수많은 레시피 중 가장 맛있는 조합을 자동 추천해주는 레시피 셰프"
H2O.ai AutoML	빠른 처리 속도 기업용 기능 지원	"대량 주문도 척척 처리하는 호텔급 자동 요리 시스템"
Microsoft AutoML	Azure 기반 AutoML 파이프라인 자동화 강점	"재료 준비부터 요리, plating까지 셰프가 전 과정 대신 해주는 풀코스 자동화 서비스"

MLOps

MLOps(Machine Learning Operations)

MLOps(Machine Learning Operations)

- 머신러닝 모델의 개발, 배포, 운영, 모니터링까지 전과정을 자동화하고 체계적으로 관리하는 방법론
- 일반 소프트웨어 개발에 DevOps가 있다면, AI 모델 개발 전체 라이프사이클에 적용된 버전
- 지속적으로 개선·관리하며 안정적으로 서비스에 활용되도록 하는 운영 체계 ➔ MLOps

MLOps의 목표 및 필요성

MLOps의 목표

- 모델 개발 속도를 높이고, 재현성과 품질을 보장
- 데이터와 모델 변경 이력을 체계적으로 관리
- 모델이 실제 서비스 환경에서 안정적으로 동작하도록 배포·모니터링
- 시간에 따라 모델 성능 저하(모델 드리프트)를 감지하고 개선할 수 있도록 자동화

필요성

- 한 번 만든 머신러닝 모델이 시간이 지나도 실제 서비스 환경에서 안정적으로 작동하도록 유지
  - 초기에 높은 성능을 보이더라도, 서비스 이후에는 데이터/사용자 행동이 변하면서 성능 하락
  - 데이터 Drift(데이터 분포 변화)가 발생하면 ➔ 예측 정확도가 낮아짐
  - 개발·배포·관리 과정을 자동화하여 불필요한 반복 업무를 줄여줌

Generative AI Trends

생성형 AI란?

생성형 AI(Generative AI)

- 많은 데이터를 학습한 후, 그 내용을 바탕으로 새로운 콘텐츠를 직접 만들어내는 인공지능
  - 단순히 기존 정보를 찾아서 알려주는 것이 아니라, 새로운 글, 이미지, 음성, 코드, 영상 등을 스스로 창작해내는 능력을 가진 AI.
- "강아지가 피아노 치는 모습을 그려줘"라고 요청
  - 문장을 이를 이해하고 새로운 그림 생성
- "친구에게 보내는 생일 축하 메시지를 감성적으로 써줘"라고 요청
  - 적절한 문장을 직접 생성
- 음악을 만들고, 사람 목소리와 비슷한 음성을 합성하고, 영상 생성 등 다양한 창작 작업에 활용

생성형 AI와 데이터사이언스와의 관계

- 데이터 사이언스는 생성형 AI의 기반 기술
  - 생성형 AI는 대규모 데이터 분석, 모델 학습, 통계·수학적 기법 등 데이터 사이언스의 토대를 바탕으로 발전
- 데이터 품질이 생성형 AI 성능에 직접 영향
  - 생성형 AI가 만드는 콘텐츠의 질은 학습한 데이터의 품질과 다양성에 따라 크게 달라짐
- 생성형 AI는 데이터 사이언스를 확장
  - 보고서 작성, 코드 생성, 데이터 시각화, 자동 모델링 등 데이터 사이언스 업무 자체를 자동화

주요 기술 트렌드

순번	기술	설명	대표 사이트
1	LLM (Large Language Model)	방대한 텍스트 학습으로 자연어 이해·요약·질의응답·생성 수행	OpenAI
2	Diffusion Model	노이즈 제거 과정을 반복하여 고해상도 이미지 생성	Stability AI
3	Multi-Modal 모델	텍스트·이미지·음성·영상 등 여러 모달리티를 동시에 처리	Google Gemini
4	AI Agents (자율 AI 에이전트)	스스로 계획하고 도구를 활용해 작업을 실행하는 지능형 에이전트	Cursor AI
5	RAG (Retrieval-Augmented Generation)	외부 지식 검색을 결합해 정확하고 신뢰도 높은 응답 생성	Perplexity
6	Video Generation AI	텍스트 기반으로 사람·사물·장면이 포함된 자연스러운 영상 생성	OpenAI Sora
7	Audio/TTS Generation	실제 사람과 유사한 음성 생성 및 음악·효과음 합성	ElevenLabs
8	3D/Simulation Model	3D 객체 생성 및 물리·환경 시뮬레이션 구현	NVIDIA Omniverse
9	Code Generation AI	자연어 요구를 기반으로 코드 생성 및 자동 개발 보조 수행	GitHub Copilot
10	Bio/Science Foundation Models	신약 개발·단백질 구조 예측 등 과학 연구를 위한 AI 모델	DeepMind Research
11	Robotics + VLM	로봇 조작, 시각 인식, 행동 계획을 VLM(비전-언어 모델)과 결합	Figure AI
12	Synthetic Data & Digital Twin Model	데이터를 AI가 가상 생성하여 학습·시뮬레이션에 활용	NVIDIA AI

수고하셨습니다 ..^^..