Pebblous

AADS LLM 파인튜닝용 QA 데이터셋 구축: 규제와 거버넌스 (EU AI Act)

(AADS LLM Fine-tuning QA Dataset Construction: Regulation and Governance Domain Report - EU AI Act)

- 작성: Pebblous (페블러스)

- 날짜: 2024. 12. 2

- 프로젝트: Pebblous AADS (Agentic AI Data Scientist)
- 주제: 규제와 거버넌스 (EU 인공지능법 중심)

I. 서론: 보고서 목적 및 EU AI Act의 전략적 중요성

본 보고서는 **AADS (Agentic AI Data Scientist) LLM**의 파인튜닝(Fine-tuning)을 위한 핵심 학습 자료 중 하나인 **유럽연합 인공지능법(EU AI Act)** 관련 QA 데이터셋의 전략적 구성과 상세 내용을 기술합니다. AADS 프로젝트의 궁극적인 목표는 데이터 과학자의 작업을 효과적으로 보조할 수 있는 **자율 AI 에이전트**를 개발하는 것입니다.

1. AADS-LLM의 목표와 AI Act의 역할

AADS LLM은 데이터 과학 이론과 규제 지식을 주입하는 **학습 자료**로서 광범위한 문서를 활용하며, 이중적인 역할(Dual Identity)을 통해 데이터의 신뢰성을 보증하는 데이터 감사관 역할까지 수행하도록 설계되었습니다.

EU AI Act 관련 소스 문서들(법령 원문, 제안서, 영향 평가, 해설 자료 등)은 LLM에게 다음과 같은 핵심 지식을 제공합니다:

- **신뢰성 및 윤리적 기반:** EU AI Act는 EU의 가치와 기본권에 기반을 두고 있으며 AI가 인간 중심적(human-centric)이며 사회에 선한 영향력(force for good)을 미치는 도구가 되도록 하는 법적 프레임워크를 제공합니다.
- **규제 지식 주입:** AI Act는 EU 단일 시장의 적절한 기능을 보장하며, **안전하고 신뢰할 수 있는 AI**의 개발 및 사용 조건을 조성하는 일반적인 목표를 달성하도록 합니다.

2. EU AI Act의 이중 목표

EU AI Act의 제안 배경에는 유럽을 디지털 시대를 위한 준비된 상태로 만들고자 하는 유럽 집행위원회

의 의제가 있습니다. 이 법안은 **'탁월성의 생태계(ecosystem of excellence)'** 와 **'신뢰의 생태계(ecosystem of trust)'** 라는 두 가지 목표를 달성하는 것을 목표로 합니다.

일반 목표 (GENERAL OBJECTIVE)	특정 목표 (SPECIFIC OBJECTIVES)
신뢰할 수 있는 AI의 개발 및 사용을 위한 조건을 조성하여 **단일 시장**의 적절한 기능을 보장합니다.	1. AI 시스템의 **안전성**과 **기본권 및 연합 가치 존중**을 보장합니다.
	2. AI에 대한 투자와 혁신을 촉진하기 위한 **법적 확실성**을 보장합니다.
	3. 기본권 및 안전 요구사항에 대한 **거버넌스 및 효과적인 시행**을 강화합니다,.
	4. 합법적이고 안전하며 신뢰할 수 있는 AI 애플리케이션의 **단일 시장 개발을 촉진**하고 시장 분열을 방지합니다,.

II. QA 데이터셋 구축 전략 및 방법론

1. 4가지 핵심 질의 유형의 적용

AADS LLM은 단순 정보 추출을 넘어, 데이터의 가치와 품질을 법적 맥락에서 판단할 수 있도록 4가지 유형의 질의-응답 쌍을 활용하여 파인튜닝 됩니다.

유형 코드	유형명	정의 (학습 목표)
A	도메인 정의/목적	AI Act의 **핵심 목적, 위험 분류 및 금지 영역**에 대한 이해를 확인합니다,.
B	데이터 구조/구성	고위험 AI 및 GPAI 모델에 필요한 **데이터 품질 기준, 데이터 거버넌스, 로깅 및 기록 유지** 의무를 확인합니다.
C	AI 모델/임무	AI 시스템의 **분류 기준, 기술적 요구사항, 적합성 평가 절차** 등 AI 기술 적용 전략에 관한 질문입니다.
D	품질/공정 관리	**사전 적합성 평가, 사후 감독, 품질 관리 시스템, 투명성 의무** 등 준수 및 집행 절차에 관한 질문입니다.

2. AI Act의 위험 기반 계층적 접근법

AI Act의 핵심은 AI 시스템을 그 잠재적 위험 수준에 따라 **계층적으로 분류**하고, 각 계층에 다른 규제 의무를 부과하는 위험 기반 접근 방식(risk-based approach)입니다.

- **수용 불가능한 위험 (Unacceptable Risk):** 안전, 생계, 기본권에 명백한 위협이 되는 AI 시스템은 금지됩니다,.
- **고위험 (High Risk):** 시장 출시 전 **엄격한 의무 및 적합성 평가(Conformity Assessment)**를 적용하며, 이는 안전 구성요소로 사용되는 AI나 기본권에 중대한 영향을 미치는 AI를 포함합니다.
- **제한적 위험 (Limited Risk):** 사용자에게 **투명성 의무**를 부과합니다 (예: 챗봇, 딥페이크).
- **최소 위험 (Minimal Risk):** 기본적으로 추가적인 규제 없이 허용됩니다 (예: 스팸 필터, 추천 시스템).

III. EU AI Act QA 데이터셋: 목표 기반 학습 자료

1. 일반 목표 (General Objective): 단일 시장 및 신뢰할 수 있는 AI 조성

유형 코드	유형명	질문 (Question)	답변 (Answer)
A	도메인 정의/목적	EU AI Act의 일반적인 목표는 무엇이며, 이는 유럽연합의 어떤 핵심 가치를 기반으로 합니까?	개입의 일반적인 목표는 **단일 시장**의 적절한 기능을 보장하고, AI 시스템이 안전하며 기존 법률과 **연합 가치를 준수**하도록 조건을 조성하는 것입니다.
B	데이터 구조/구성	AI Act가 안전하고 신뢰할 수 있는 AI를 목표로 할 때, 데이터의 획득 및 사용과 관련하여 요구되는 정보 의무(IOs)는 무엇입니까?	데이터 획득에 관한 정보 의무에는 데이터 수집 과정, 데이터의 출처, 그리고 개인정보의 경우 원래의 수집 목적에 대한 상세 내용을 제공하는 것이 포함됩니다.
C	AI 모델/	AI Act가 추구하는 '신뢰할 수 있는 AI'를 입증하기 위해 제공자가 기술 문서에 명시해야 하는 모델의 성능 특성	AI 시스템의 성능은 의도된 목적을 달성하는 데 있어 **정확성, 공정성, 견고성 및 안전성**의 예상되는 수준과 오차 범위가 기술되

	임무	세 가지는 무엇입니까?	어야 합니다.
D	품질/공정 관리	고위험으로 분류되지 않은 AI 시스템의 신뢰성을 높이기 위해 AI Act가 권고하는 자발적 메커니즘은 무엇이며, 이에 대한 평가 주기는 어떻게 됩니까?	**자발적 행동 규범(voluntary codes of conducts)** 의 영향과 효과를 평가해야 하며, 이는 **발효일로부터 4년 후**부터 3년마다 이루어져야 합니다.

2. 특정 목표 1 (Specific Objective 1): 안전성 및 기본권 존중

유형 코드	유형명	질문 (Question)	답변 (Answer)
A	도메인 정의/목적	첫 번째 특정 목표가 고위험 AI 시스템에 대해 달성하고자 하는 두 가지 핵심 보장 요소는 무엇입니까?	AI 시스템이 시장에 출시되거나 사용될 때 **안전성(safety)** 을 보장하고, 기존 법률상의 **기본권 및 연합 가치**를 존중하도록 하는 것입니다.
B	데이터 구조/구성	고위험 AI 시스템 제공자는 시스템이 기본권에 미칠 수 있는 영향을 줄이기 위해 기술 문서에 어떤 종류의 위험을 명시해야 합니까?	제공자는 AI 시스템이 제기하는 **잠재적 부작용(side effects)** 과 **기본권 위험**을 문서화해야 하며, 이는 시스템의 정확성, 공정성, 견고성 및 안전성에 영향을 미칠 수 있는 모든 예측 가능한 상황을 포함합니다.
C	AI 모델/임무	교육 및 직업 훈련 기관에서 고위험 AI 시스템으로 분류되는 두 가지 구체적인 사용 사례를 제시하십시오.	개인이 받을 수 있거나 접근할 수 있는 **교육 수준을 평가**하는 목적으로 사용되는 AI 시스템, 또는 시험 중 학생의 **금지된 행동을 모니터링 및 탐지**하는 목적으로 사용되는 AI 시스템입니다.
D	품질/공정	고위험 AI 시스템 제공자가 시스템의 지속적인 규정 준수를 보장하기 위해 시장 출시 전에 갖추어야 하는 공식적인 관리 시스템은	시스템의 라이프사이클 전반에 걸쳐 규정 준수를 보장하는 **품질 관리 시스템(Quality Management System)** 을 구축하고 이행해야

	관리	무엇입니까?	합니다.

3. 특정 목표 2 (Specific Objective 2): 법적 확실성 확보

유형코드	유형명	질문 (Question)	답변 (Answer)
A	도메인 정의/목적	두 번째 특정 목표인 '법적 확실성 확보'는 AI 분야의 투자 및 혁신에 어떤 긍정적인 영향을 미치도록 의도됩니까?	명확하고 조화된 법적 프레임워크를 제공함으로써 시장 불확실성을 줄이고, AI 기술의 개발 및 배포를 위한 **투자**와 **혁신**을 용이하게 합니다.
B	데이터 구조/구성	법적 확실성을 위해 고위험 AI 시스템 제공자는 AI 시스템 개발에 사용된 리소스와 관련된 어떤 상세 정보를 문서화해야 합니까?	제공자는 AI 시스템 개발, 훈련, 테스트 및 검증에 사용된 **계산 자원(computational resources)** 과 함께, 데이터를 **표준화된 프로토콜**에 따라 처리했음을 증명하는 문서를 갖춰야 합니다.
C	AI 모델/임무	AI Act의 규제 요구사항 준수를 위한 기술적 해결책과 관련하여 제공자가 반드시 문서화해야 하는 중요한 결정 사항은 무엇입니까?	제3장 제2절에 명시된 요구사항을 준수하기 위해 채택된 기술적 해결책과 관련하여 이루어진 **가능한 모든 상충 관계(trade-off)** 에 대한 결정을 문서화해야 합니다.
D	품질/공정 관리	AI 혁신을 장려하면서도 규제 감독을 제공하기 위해 마련된 통제된 환경은 무엇이며, 이를 통해 무엇을 향상시키고자 합니까?	**인공지능 규제 샌드박스**가 설립되며, 이는 규제 감독 하에 혁신적인 AI 시스템을 개발, 훈련, 검증 및 테스트할 기회를 제공하여 **법적 확실성**을 향상시키고자 합니다.

4. 특정 목표 3 (Specific Objective 3): 거버넌스 및 효과적인 집행 강화

유	유		

형코드	형명	질문 (Question)	답변 (Answer)
A	도메인 정의/목적	세 번째 특정 목표인 거버넌스 강화는 무엇의 효과적인 집행을 보장하는 데 중점을 둡니까?	AI 시스템에 적용 가능한 **기본권 및 안전 요구사항**의 거버넌스 및 **효과적인 집행**을 강화하는 데 중점을 둡니다.
B	데이터 구조/구성	규제 집행을 지원하기 위해 설립된 **과학 패널(Scientific Panel)** 은 GPAI 모델 평가와 관련하여 어떤 데이터 관련 기능을 수행합니까?	과학 패널은 범용 AI 모델 및 시스템의 역량을 평가하기 위한 **도구 및 방법론 개발**에 기여하며, 이는 **벤치마크 (benchmarks)** 개발을 포함합니다.
C	AI 모델/임무	고위험 AI 시스템의 시장 출시 전에 규정 준수를 보장하기 위해 회원국이 지정해야 하는 핵심적인 거버넌스 주체는 무엇이며, 이들의 역할은 무엇입니까?	각 회원국은 적합성 평가 기관의 평가, 지정 및 통보에 필요한 절차를 수립하고 수행할 책임을 지는 **통보 당국(Notifying Authority)** 을 지정해야 합니다.
D	품질/공정관리	거버넌스 시스템의 일환으로, **자문 포럼 (Advisory Forum)** 의 활동 결과는 어떤 방식으로 공개되어야 합니까?	자문 포럼은 활동에 대한 **연례 보고서**를 작성해야 하며, 이 보고서는 **대중에게 공개**되어야 합니다.

5. 특정 목표 4 (Specific Objective 4): 단일 시장 개발 촉진 및 시장 분열 방지

유형코드	유형명	질문 (Question)	답변 (Answer)
	도메인	네 번째 특정 목표가 달성하고자 하는 경제적 이점은 무엇이며, 규제적 관점	합법적이고 안전하며 신뢰할 수 있는 AI 애플

A	정의/목적	에서 방지하고자 하는 부정적 결과는 무엇입니까?	리케이션의 **단일 시장 개발**을 촉진하고, 회원국 간의 상이한 규정으로 인한 **시장 분열 (market fragmentation)** 을 방지합니다.
B	데이터 구조/구성	AI 사무국이 위험 수준 평가 방법론을 개발할 때, 시장 분열 방지 및 규제 일관성을 위해 고려해야 하는 세 가지 규제 목록 기준은 무엇입니까?	**Annex III 목록(고위험 시스템)**, **금지 관행 목록(Article 5)**, 그리고 **추가적인 투명성 조치가 필요한 AI 시스템 목록(Article 52)** 에 새로운 시스템을 포함하는 기준을 바탕으로 합니다.
C	AI 모델/임무	영향 평가 보고서에 따르면, AI Act의 규제 준수 비용과 관련하여 2020년 가장 높은 비용 가중치(0.210)를 차지할 것으로 예상된 산업 부문은 어디였습니까?	**제조업(Manufacturing)** 이 2020년에 0.210의 가중치로 가장 높은 규제 준수 비용 가중치를 차지할 것으로 예상되었습니다.
D	품질/공정 관리	통보 당국(Notifying Authorities)은 시장 분열을 방지하고 집행의 통일성을 확보하기 위해 적합성 평가 절차를 어떻게 개발해야 합니까?	적합성 평가 기관의 평가, 지정 및 통보 절차는 **모든 회원국의 통보 당국 간의 협력**을 통해 개발되어야 합니다.

IV. 결론 및 AADS 프로젝트 기여

1. 규제 및 거버넌스 지식의 체계화

EU AI Act 관련 QA 데이터셋은 AADS LLM이 데이터 과학자로서의 역량뿐만 아니라, **글로벌 규제 환경** 내에서 AI 시스템의 신뢰성과 안전성을 객관적으로 평가할 수 있는 **책임성 있는 에이전트**로서 기능하도록 돕습니다.

이 데이터셋은 AI Act가 **위험도에 따라 규제 의무를 차별화**하는 구조(금지, 고위험, 제한적 위험)를 LLM이 명확히 이해하도록 파인튜닝 합니다,. 특히, 고위험 AI 시스템에 대한 **데이터 품질(High Quality Datasets) 및 거버넌스 요구사항**,에 대한 심층적인 질의응답은 AADS LLM이 데이터 감사관(Data Auditor) 역할을 수행할 수 있는 논리적 기반을 제공합니다.

2. LLM 파인튜닝의 기반 강화

규제와 거버넌스 도메인 특화 데이터셋은 LLM에게 단순 법조문 해석을 넘어, 규제가 탄생하게 된 **정책적 맥락과 목적** (도메인 정의/목적), **AI 모델의 구조적 요건** (AI 모델/임무), 그리고 **법적 준수를 위한 데이터 라이프사이클 관리 절차** (품질/공정 관리)에 대한 이해를 주입합니다. 이는 **데이터 품질과 신뢰성이 규제 준수의 핵심**임을 AADS LLM이 추론할 수 있도록 합니다.

이처럼 EU AI Act QA 데이터셋은 AADS LLM이 복잡하고 빠르게 변화하는 규제 환경에서 **법적 리스크를 선제적으로 식별**하고, 고객에게 **AI Ready 데이터**를 보증할 수 있는 핵심 지능적 자산이 될 것입니다.

Pebblous

> Pebblous Makes Data Tangible

contact@pebblous.ai