유럽연합

EU 인공지능법 최종 확정 : 2021년 초안과 무엇이 달라졌나

임현철 전문연구원 (한국행정연구원 규제정책연구실)

2024년 2월 2일, 「EU 인공지능 법(EU Artificial Intelligence Act)」이 최종 합의에 이르렀다. 2021년 4월 유럽연합 집행위원회가 최초 발의한 이래, 유럽 집행위원회, 유럽 의회, 그리고 유럽연합 회원국들의 복잡한 협상과 수정 과정을 거쳐, 마침내 합의에 도달했다. 최종안은 1)AI의 안정성 확보, 2)AI 산업 혁신 및 투자확대, 3)AI 관련 기관의 효과적인 거버넌스 및 집행력 확보, 4)시장분열 방지라는 법안의 목적 등에 있어서 2021년 유럽연합 집행위원회가 발의한 최초의 「EU 인공지능 법안」과 같지만, 세부내용에서는 상이한 부분이 있다. 본고에서는 최종안이 초안에 비해서 어떠한 점이 달라졌는지를 살펴보고자 한다. 초안과 최종안의 주된 차이점은 아래 〈표 1〉과 같이 정리될 수 있다.

〈표 1〉 「EU 인공지능 법안」 초안과 최종안의 차이

구분	2021년 초안	2024년 최종안
AI의 정의	"기계학습, 논리/지식 기반 접근, 통계적 접근, 베이즈 추정, 검색 및 최적화 방법을 활용하여 개발된 소프트웨어로, 인간이 정의한 일련의 목적을 위해 콘텐츠·예측·추천·결정과 같은 산출물을 생산하는 소프트웨어"	"자율성을 가지고 작동하도록 설계된 시스템으로, 적응력을 발휘할 수 있으며, 입력값을 받아 명시적 또는 암묵적 추론을 통해 환경에 영향을 미칠 수 있는 콘텐츠·예측·추천·결정과 같은 산출물을 생산하는 소프트웨어"
적용범위	○ 비전문적 영역에서 사적 목적으로 AI를 활용하는 사용자는 법안의 대상에서 제외	○ 비전문적 영역에서 사적 목적으로 AI를 활용하는 사용자는 법안의 대상에 포함
범용 AI (생성형AI)	○ 별도 규정 없음	○ 규모에 따른 차등 규제
벌금	○ 고위험: 3,000만 유로 또는 전년도 매출의 6% ○ 의무위반: 2,000만 유로 또는 전년도 매출의 4% ○ 유해정보: 1,000만 유로 또는 전년도 매출의 2%	○ 고위험: 3,500만유로 또는 전년도 매출의 7% ○ 의무위반: 1,500만 유로 또는 전년도 매출의 3% ○ 유해정보: 750만 유로 또는 전년도 매출의 1.5%
허용불가 AI	○ 인간의 안전, 보안 및 기본권에 위험을 초래할 수 있는 AI	○ 인간의 안전, 보안 및 기본권에 위험을 초래할 수 있는 AI ○ 생체분류 AI, 안면인식 AI 금지조항 등 추가
AI 생성 콘텐츠 표시	○ 별도 규정 없음	○ 의무화

출처: 저자 작성

국제표준을 목표로 OECD의 AI 정의 차용

AI의 정의가 달라졌다. 기존 안에서는 AI 시스템을 "기계학습, 논리/지식 기반 접근, 통계적 접근, 베이즈 추정, 검색 및 최적화 방법을 활용하여 개발된 소프트웨어로, 인간이 정의한 일련의 목적을 위해 콘텐츠·예측·추천·결정과 같은 산출물을 생산하는 소프트웨어"로 정의했다. 하지만 최종안에서는 AI 시스템을 "자율성을 가지고 작동하도록 설계된 시스템으로, 적응력을 발휘할 수 있으며, 입력값을 받아 명시적 또는 암묵적 추론을 통해 환경에 영향을 미칠 수 있는 콘텐츠·예측·추천·결정과 같은 산출물을 생산하는 소프트웨어"로 정의한다. 이는 OECD의 AI 시스템에 대한 정의[1]를 차용한 것이다. 이 같이 AI에 대한 정의가 OECD의 정의로 변경된 이유는, 전통적인 소프트웨어와 AI 소프트웨어를 구분짓는 결정적 차이를 "추론"이라고 강조하는 의미도 있지만, 「EU 인공지능 법」의 국제적 연계를 고려하고 있기 때문으로 보인다(EY, 2024).[2]

파급력이 강한 범용 AI(생성형 AI)에 추가의무 부과

「EU 인공지능 법」의 적용 범위가 확대되었다. 최초 법안에서는 비전문적 영역에서 사적 목적으로 AI를 활용하는 사용자는 법안의 대상에서 제외되었으나, 최종안에서는 이들도 적용대상에 포함되었다. 또한 최종안에서는 사회적 파급력이 크다는 이유로, 범용 인공지능(General Purpose AI: GPAI)을 별도의 장(Chapter)[3]으로 다루며 추가의무를 부과하고 있다. 이 장에서는 GPAI를 "다량의 데이터를 사용하여 훈련되고 자가진단이 가능한 AI이며, 단독으로 사용될 수도 있고 다른 AI 시스템과 통합하여 다양한 목적을 수행할 수 있는 소프트웨어"라고 정의하고 있다. GPAI는 규모에 따라 차등규제를 적용받는데, GPAI는 훈련에 사용되는 계산량이 10^{25}FLOPS[4]가 넘으면 영향력이 강한(systemic) AI라 판단하여[5] 지켜야 할 의무가 증가된다. GPAI 공급자가 지켜야 할 규제를 간략히 살펴보면 아래 〈표 2〉과 같다.

---

1) https://www.oecd-ilibrary.org/docserver/623da898-en.pdf?expires=1710134702&id=id&accname=ocid53022015&checksum=CB0FC4D59A0EC77C8FF55A68D6D46C6E (검색일: 2024. 3. 11.)
2) https://assets.ey.com/content/dam/ey-sites/ey-com/en_gl/topics/ai/ey-eu-ai-act-political-agreement-overview-february-2024.pdf (검색일: 2024. 3. 11.)
3) TITLE VIIIA, GENERAL PURPOSE AI MODELS
4) 1초에 수행할 수 있는 부동 소수점 연산의 횟수
5) GPAI 제공업체는 GPAI가 위 기준을 충족하는 경우, 2주 이내에 위원회에 한다.

〈표 2〉 GPAI 공급자가 지켜야 할 규제

Base level tier GPAI	Systemic risk tier GPAI
○ GPAI 최신 기술문서 유지 (트레이닝, 테스트, 평가결과 포함) ○ GPAI의 기능과 한계에 대한 정보 제공 ○ 저작권 준수를 위한 지침 수립 ○ GPAI의 트레이닝에 사용한 데이터 공개	○ GPAI 기술문서 작성 (트레이닝, 테스트, 평가결과 포함) ○ GPAI의 기능과 한계에 대한 정보 제공 ○ 저작권 준수를 위한 지침 수립 ○ GPAI의 트레이닝에 사용한 데이터 공개 ○ 위험식별 및 완화를 위한 모델평가 수행 ○ AI 시스템의 위험 평가 및 완화 ○ 모델에 적대적 테스트[6] 수행 ○ 심각한 문제 발생 시, 문제를 추적 및 문서화하고, 지체없이 관계당국에 보고 ○ 사이버 보안, 물리적 보호 확인 ○ 모델의 예상 에너지소비량 보고

출처: 저자 작성

고위험 분야의 벌금은 더 높게, 그 외 분야는 더 낮게

의무위반 시 부과되는 벌금은 일부는 강화되고 일부는 약화되었다. 위험도가 높은 AI가 의무사항을 위반한 경우 부과되는 벌금은 초안에서는 3,000만 유로 또는 전년도 총 매출액의 6% 중 높은 액수였다. 하지만 최종안에서는 3,500만 유로 또는 전년도 총 매출액의 7% 중 높은 액수로 상향조정되었다. 반면 기타요건 및 의무의반 시 부과되는 벌금은 다소 낮아졌는데, 초안은 2,000만 유로 또는 전년도 총 매출액의 4% 중 높은 액수 였으나, 최종안은 1,500만 유로 또는 전년도 총 매출액의 3% 중 높은 액수로 규정하고 있다. 잘못된 정보를 제공 시 지불해야하는 벌금도 초안은 1,000만 유로 또는 전년도 총 매출액의 2%로 규정하고 있지만, 최종안에서는 750만 유로 또는 전세계 매출액 1.5%로 규정하여 다소 완화된 것을 확인할 수 있다.

유럽 인공지능 법의 규제 체계 : 생체분류·안면인식 AI 금지조항 등의 추가

초안과 마찬가지로 최종합의된 「EU 인공지능 법」에서도, AI를 위험에 따라 분류하여 달리 취급하는 위험 기반 접근(risk-based approach)을 취하고 있는 것은 변함없다. 위험 기준 또한 1)허용할 수 없는 위험(unacceptable risk), 2)고위험(high risk), 3)제한적 위험(limited risk), 4)저위험(minimal risk)으로 구분하여 차등규제하는 것 또한 2021년 초안과 같다. 하지만 개인의 특징을 활용하는 생체분류시스템 금지, 안면인식 데이터베이스 구축을 위한 CCTV 영상의 얼굴 캡쳐 금지 등 내용측면에서는 변경사항이 있다. 아래 〈표 3〉는 「유럽 인공지능 법」의 개략적인 체계이다. GPAI는 아래 체계도에 따르지 않고 별도로 관리된다.

6) 적대적 프롬프팅은 사용자가 악의를 가지고 프롬프트를 주입하여 목표를 변환하려 하거나, 기밀 정보 등을 유출하는 것을 의미한다.

〈표 3〉「EU 인공지능 법」 체계

분류	금지여부	사례	설명
허용 불가 AI	인간의 안전, 보안 및 기본권에 심대한 위험을 초래할 가능성이 높아 금지	○ 잠재의식 활용 시스템	개인의 잠재의식을 활용하여 어린이, 노약자, 장애인 등 특정그룹을 착취·조작하여 개인의 자유의지에 피해를 줄 수 있는 시스템
		○ 인간 평점 시스템	공공기관이 인간의 행동과 인격적 특성을 활용해 개인을 평가하여 분류하는 AI 시스템
		○ 감정 인식 시스템	직장이나 학교에서 인간의 감정을 추론해 평가에 활용하는 시스템. 파일럿의 졸음 감지 등 일부 상황에서는 예외적으로 허용
		○ 생체분류 시스템	개인의 육체·정신적 특성을 활용하여 정치성향·종교신념·인종·성적취향 등 민감한 데이터를 추론하기 위한 시스템. 사법기관에서는 사전허가를 받으면 중대한 범죄를 저지른 자에 한해서 생체분류시스템 허용
		○ CCTV 얼굴 캡쳐 시스템	얼굴인식 데이터베이스를 획득하기 위해 인터넷이나 CCTV에서 개인의 얼굴이미지를 무차별적으로 스크랩 하는 시스템
		○ 개인행동 예측 시스템	개인의 범행·재범행 가능성과 같은 개인행동을 예측하는 시스템
고위험 AI	「EU 인공지능 법안」의 의무사항을 준수하면 허용	○ 안전관련 시스템	의료장비, 자동차, 기계장치, 민항기, 해양용품, 농업차량, 철도용품, 장난감 등의 안전장치로 활용되는 AI 시스템
		○ 생체인증·분류 시스템	생체정보기반 실시간 개인 식별 시스템 등
		○ 기반시설 운영 시스템	도로·수도·가스·전력의 운영·관리 시스템 등
		○ 공공서비스 관련 시스템	공공서비스 수급자격 평가, 긴급출동서비스 우선권 설정 시스템 등
		○ 교육·직업훈련 시스템	학생 평가 시스템 등
		○ 채용·직원관리 시스템	채용·승진·해고, 직무배분, 성과평가, 행동감시 시스템 등
		○ 법집행 시스템	범죄증거 평가를 위한 데이터분석 시스템 등
		○ 이민·난민·국경관리 시스템	입국자에 대한 리스크 평가 및 거짓말 탐지, 위조서류 적발 시스템 등
		○ 사법행정·민주절차 시스템	분쟁해결을 위한 사실 조사·해석, 선거관리 시스템 등
제한적 위험 AI	투명한 정보공개의 의무를 이행하면 허용	○ 사람과 상호작용 하는 시스템	챗봇 등
		○ 딥페이크 시스템	AI에 의해 시각적, 청각적으로 조작된 콘텐츠
저위험 AI	별다른 의무 없이 허용	○ 위 범주에 포함되지 않는 AI 시스템	
GPAI	○ 별도관리		

출처: 저자 작성

위 〈표 3〉에서 보듯이, 고위험과 저위험 AI 소프트웨어를 개발·발매·유통·활용하려면 「EU 인공지능 법」에서 규정한 요건을 충족해야 한다. 고위험 AI 시스템은 데이터 거버넌스, 정보공개, AI DB 등록, 사이버보안 등의 조건을 갖춰야 한다. 이를 위해 「EU 인공지능 법」은 위험 관리 시스템 구축, 사람에 의한 관리·감독, 기술문서화 및 기록보존, 관계기관과의 협력 등을 요구한다. 즉, 이러한 규제를 통하여, EU의 AI 법안은 AI 시스템이 인간의 안전, 보안, 기본권을 침해하는지를 추적, 관찰, 수정할 수 있는 생태계를 마련하려는 것이다. 제한된 위험의 AI의 경우에는 이용자에게 AI와 상호작용하고 있다는 공개 또는 고지의 의무만 부과된다. 마지막으로 저위험 AI의 경우, 법적 효력이 있는 규제로 제한되지 않으며, 사업자의 자율규제에 맡긴다. 아래 〈표 4〉는 「EU 인공지능 법」의 규제 체계를 정리·요약한 자료이다.

〈표 4〉 「EU 인공지능 법안」의 규제 체계

구분	규제	내용
허용할 수 없는 AI	○ 허용 불가	
고위험 AI	위험관리 시스템 구축	○ AI 시스템의 전 생애주기에 걸친 위험관리시스템을 마련하고 정기적으로 최신화 - 예상가능한 리스크 식별·분석 - 적합한 위험 관리 방안의 채택 - 출시 전, 자체 또는 독립평가자를 통해 규정 준수 입증 - 출시 후, 모니터링을 통한 발생가능한 위험 추산·검토 - 아동에 미치는 위험 고려
	효율적인 데이터 거버넌스	○ 트레이닝·검증 데이터의 대표성·관련성·완전성 - 데이터 수집·전처리 방안 마련 - 데이터셋의 규모, 적합성 사전 평가 - 데이터의 편향 검토 - 데이터의 오류 식별 방안 마련 ○ 민감정보, 고유식별정보, 전과 등을 처리할 경우 합당한 보안조치
	문서화 및 기록보존	○ AI 시스템에서 사용한 기술을 문서로 기록하고 지속적으로 업데이트 - 시스템 개발과정 - 기능·제어·감독 방법 - 위험 관리 방법 - 모델 성능평가 내용 ○ 문서는 규제당국이 고위험 AI가 규제를 준수했는지 바로 확인할 수 있을 정도로 상세히 기록 ○ 로그(이벤트 기록)가 자동적으로 기록될 수 있도록 개발
	투명한 정보공개	○ 사용자가 시스템을 이해할 수 있도록 상세한 설명서 제공 ○ 설명서를 디지털 포맷 등 다양한 방식으로 제공
	인간에 의한 감시·감독	○ 인간이 인간의 안전·보안·기본권에 대한 위험 여부를 감시할 수 있도록 설계 - 오작동·오류·예상 밖의 상황을 포착할 수 있는 인력 배치 - 오류가 발생한 상황에서 인간에 의해 시스템 중단 가능
	정확성·강건성·보안	○ 머신러닝 시스템은 출력값이 향후 입력값으로 투입되는데, 잘못된 출력값을 다시 입력하여 생기는 편향을 복원할 수 있도록 개발

		○ 미인가된 활용, 기능의 변경 시도에 대한 보안 마련 - 데이터셋을 조작하려는 사용자에 대한 보안
	고위험 AI DB에 등록	○ EU 데이터베이스에 고위험 AI 시스템을 등록할 의무
제한적 위험 AI	정보공개	○ 인간과 AI가 상호작용중이고 표시 ○ 딥페이크는 콘텐츠가 인공적으로 생성·조작되었다고 고지
저위험 AI	○ 별도 의무 없음	

출처: 고학수 외(2021)을 토대로 재구성[7)]

규제샌드박스 제도의 의무화

한편, 「EU 인공지능 법」은 AI 규제에 대한 내용뿐만 아니라, AI 혁신을 지원하는 내용도 포함하고 있다. 「EU 인공지능 법」은 회원국의 AI 혁신을 지원하기 위해, AI 규제샌드박스 제도의 구축을 의무화하고 있다. 규제샌드박스는 "사업자가 신기술을 활용한 새로운 제품과 서비스를 일정 조건(기간·장소·규모 제한)하에서 시장에 우선 출시해 시험·검증할 수 있도록 현행 규제의 전부나 일부를 적용하지 않는 제도"이다[8)]. 「EU 인공지능 법」은 이 같은 규제샌드박스 제도를 각 회원국이 독자적으로 운영하는 것이 아니라, 유럽 전역에서 일관되게 사용할 수 있는 공통의 규칙으로 만들어야 한다고 명시하고 있다. 유럽연합은 회원국 전역에 걸쳐 공통으로 운영되는 규제샌드박스를 통해 AI 시스템의 시장출시를 앞당길 계획이며, 또한 놓치고 있거나, 과도하다고 판단되는 규제를 파악하여 「EU 인공지능 법」의 품질향상을 도모할 예정이다.

유럽 인공지능 법의 향후 추진 일정

「EU 인공지능 법」은 2024년 2~3분기에 유럽의회와 위원회의 승인을 거쳐 발효될 예정이다. 「EU 인공지능 법」은 지침이 아닌 규정으로, 개별 국가의 법률과 상관없이 회원국에 직접 효력이 발생한다. 법안이 효력을 발휘하면, EU집행위는 AI사무국(EU 감독기관)을 설립해야 하며, 회원국은 AI 규제샌드박스 제도를 위한 규정을 의무적으로 마련해야 한다. 법안 발효 후 6개월 후인 2024년 4분기에서 2025년 1분기 부터는 「EU 인공지능 법」의 금지조항의(Prohibitions) 효력이 생길 예정이다. 2025년 2분기~3분기에는 GPAI 시스템에 대한 규제, 2026년부터는 고위험 AI에 대한 규제가 시작될 예정이다.

7) 고학수, 임용, 박상철(2021), 유럽연합 인공지능법안의 개요 및 대응방안, DAIG 2021년 제2호
8) 규제샌드박스 포털 https://www.sandbox.go.kr/sandbox/info/sandbox_intro.jsp

참고자료

- https://artificialintelligenceact.eu/the-act/
- https://assets.ey.com/content/dam/ey-sites/ey-com/en_gl/topics/ai/ey-eu-ai-act-political-agreement-overview-february-2024.pdf
- http://sapi.co.kr/wp-content/uploads/2021/09/%EC%9C%A0%EB%9F%BD%EC%97%B0%ED%95%A9-%EC%9D%B8%EA%B3%B5%EC%A7%80%EB%8A%A5%EB%B2%95%EC%95%88%EC%9D%98-%EA%B0%9C%EC%9A%94-%EB%B0%8F-%EB%8C%80%EC%9D%91%EB%B0%A9%EC%95%88_0923_2.pdf
- https://www.oecd-ilibrary.org/docserver/623da898-en.pdf?expires=1710134702&id=id&accname=ocid53022015&checksum=CB0FC4D59A0EC77C8FF55A68D6D46C6E