유럽연합은 지난 12월 8일(현지시각) 세계 최초로 고위험 인공지능을 포괄적으로 규제하는 '인공지능법안'(AI Act)에 대해 합의하였습니다. 이 법은 유럽연합 역내에서 활용되는 인공지능에 적용되지만, 상품과 서비스가 전 세계 각국을 넘나드는 현실을 고려할 때 한국을 비롯한 각국의 업체와 규범 형성에 영향을 미칠 것으로 예상됩니다. 이에 사단법인 정보인권연구소와 진보네트워크센터는 이 법안에 대한 이해를 돕기 위해 합의안의 주요 내용과 그 함의에 대한 분석 문서를 작성하였습니다. 국내에서 인공지능 규율을 위한 법제 마련과 토론에 도움이 될 수 있기를 바랍니다.

EU 인공지능 법안(AI Act) 합의안 분석

작성 : (사)정보인권연구소, 진보네트워크센터

1. 도입

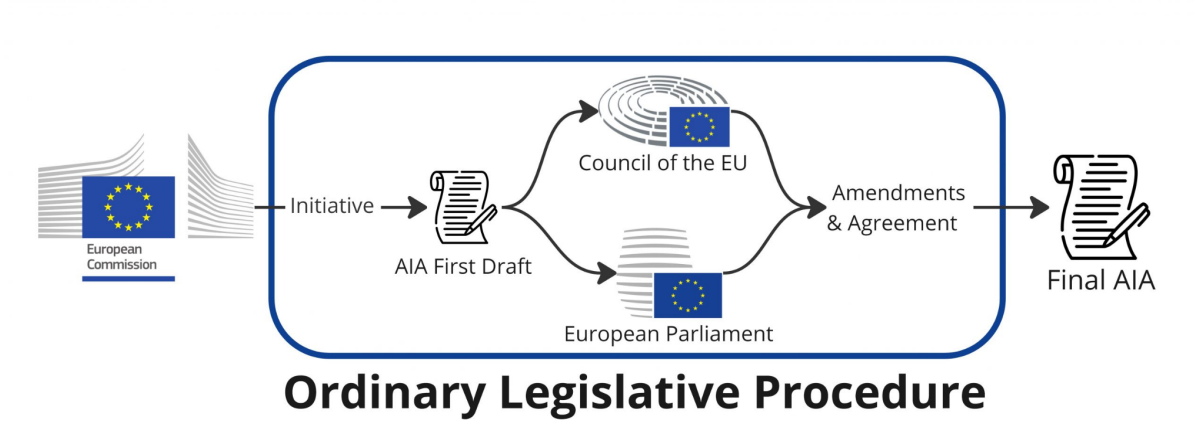
- 지난 12월 8일(현지시각), 유럽연합 이사회와 유럽의회는 '인공지능법안'(AI Act)에 대해 잠정 합의함. 이 법은 일부 예외 조항을 제외하고, 발효 후 2년 후에 시행될 예정이라고 함.
- 인공지능 기술이 빠르게 발전하고 관련 상품과 서비스가 공공 및 민간 분야에서 이미 도입되어 활용되고 있음. 그 동안 국제적인 범위에서 인공지능에 대한 규범화 노력이 이어졌지만, 구속력이 없는 규범이기 때문에 규율 내용이 추상적이고 선언적이었음. 미국에서 구속력이 있는 대통령 행정명령으로 인공지능의 안전하고 신뢰할 수 있는 개발 및 사용에 관한 행정 명령을 발표했지만, 구속력 있는 규율은 연방정부에 공급하는 인공지능 시스템에 대한 것이고, 행정부의 구속력 없는 가이드라인 및 표준 제정을 규정하고 있어서 전분야의 인공지능 규율은 아님.
- 반면, 유럽연합 인공지능 법은 세계 최초로 인공지능을 포괄적으로 규율하는 법이라는 의미를 가지고 있음. 비록 이 법은 유럽연합 역내에서 활용되는 인공지능에 적용되지만, 상품과 서비스가 전 세계 각국을 넘나들고, 특히 인공지능의 경우 생성형 인공지능이나 범용 인공지능을 비롯한 핵심 분야는 사실상 전세계를 대상으로 하지 않을 수 없다는 점에서 그 영향이 클 수 밖에 없음. 유럽연합은 개인정보보호법제(GDPR)와 디지털 플랫폼과 디지털 서비스 분야(DMA, DSA)에서 브뤼셀 효과라고 하는 전세계 규범화(Global standards) 효과를 보여준 바 있음. 이런 점을 고려한다면 이 법은 한국을 비롯한 다른 나라의 규범 형성시 참조하지 않을 수 없는 중요한 규범이 될 것으로 전망됨.

- 인공지능 규제와 관련하여 유럽연합은 인공지능 산업이 발전하지 못했기 때문에 미국의 빅테크를 규제하기 위하여 강력한 규제 체제를 만드는 것이라고 폄하하면서 기업의 자율규제를 옹호하는 주장이 있는데, 이는 기업의 이익을 옹호하는 논리에 다름 아님. 각국의 인공지능 규제 정책이 비단 자국의 산업 육성 논리에 좌우된다면, 산업 육성을 위해서 안전과 인권을 위협하는 인공지능을 허용할 수 있다는 것인지 의문임. 자국 산업 육성은 부차적인 고려사항일 뿐이며, 문제는 인공지능의 위험성을 통제할 수 있는 적절한 규제 체제를 구축하는 것임. 또한 기업들의 자율규제에 맡겨야 한다는 주장이 있는데, 과연 국내에서 기업들이 아무런 규제없이 충분한 수준의 규제 조치를 자율적으로 취한 적이 있었는지 의문임.
- 아직 합의안의 구체적인 조문은 공개되지 않아 정확한 분석을 하기에는 한계가 있지만, 인공지능 법안 초안(EC의 proposal), 유럽연합 이사회나 유럽의회의 수정제안, 그리고 잠정 합의에 대한 보도자료를 토대로 합의안의 대략적인 방향 및 그 함의를 추측해볼 수 있음.

2. 주요 경과

- 2021.4.21. 유럽 집행위원회, [유럽연합 인공지능법안(AI Act)] 초안 발표 (이하 초안)
- 2022.12.6. 유럽연합 이사회(Council of the European Union), 인공지능 법안에 대한 공동 입장 발표
- 2023.6.14. 유럽의회, 인공지능 법안에 대한 협상안 발표(이하 의회안)
- 2023.12.9. 집행위원회, 이사회, 유럽의회 3자 협상을 통해 인공지능 법안에 대해 잠정 합의.
- 향후 몇 주 동안 세부 조문 정리 작업을 거쳐, 이사회 및 유럽의회에서 공식적으로 채택이 될 예정임.

참고 : 유럽연합의 입법절차



by Hadrien Pouget

3. 인공지능 법안 합의안(이하 합의안)의 주요 내용 및 함의

(1) 인공지능의 정의

- 인공지능 시스템의 정의가 단순한 소프트웨어 시스템과 인공지능을 구별하는 데 충분히 명확한 기준을 제공하도록 하기 위해, 합의안은 OECD의 정의를 차용하여 인공지능을 정의함.
    - 초안에서는 부속서I 에서 법 적용의 대상이 되는 인공지능 기술 및 접근 방식을 나열하였으나, 의회안에서는 인공지능 분야의 최신의 기술 발전을 반영하면서 OECD의 인공지능 정의를 차용함.
    - 의회안은 다음과 같이 되어 있는바, 합의안도 이와 유사할 것으로 보임
        - “artificial intelligence system’ (AI system) means a machine-based system that is designed to operate with varying levels of autonomy and that can, for explicit or implicit objectives, generate outputs such as predictions, recommendations, or decisions, that influence physical or virtual environments
        - ‘인공 지능 시스템(Artificial Intelligence system)’(AI 시스템)은 다양한 수준의 자율성을 가지고 작동하도록 설계되어 명시적 또는 암시적 목표에 따라 물리적 또는 가상 환경에 영향을 미치는 예측, 추천 또는 결정과 같은 결과물을 생성할 수 있는 기계 기반 시스템을 의미한다.
    - OECD의 인공지능 정의
        - a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments.

        - ■ 사람이 정의한 특정 목표 집합에 대해 실제 또는 가상 환경에 영향을 미치는 예측, 권장 사항 또는 결정을 내릴 수 있는 기계 기반 시스템.
    - 의회안의 서설(recital) 수정안에 따르면 “본 규정에서 AI 시스템의 개념은 명확하게 정의되어야 하며, 법적 확실성, 조화 및 폭넓은 수용을 보장하기 위해 인공지능을 연구하는 국제기구의 작업과 밀접하게 연계되어야 하며, 동시에 본 규정의 급속한 기술 발전을 수용할 수 있는 유연성을 제공해야 합니다”(서설 6)라고 OECD 정의를 차용한 취지를 설명하고 있음.
    - 이 정의는 "단순한 소프트웨어 시스템이나 프로그래밍 접근 방식과 구별할 수 있도록 학습, 추론 또는 모델링 기능과 같은 인공지능의 주요 특성에 기반"해야 하며, "다양한 수준의 자율성을 가지고 작동하도록 설계되어 있으며, 이는 인간의 제어로부터 어느 정도 독립적으로 행동하고 인간의 개입 없이도 작동할 수 있는 기능을 갖추고 있"는 시스템을 의미함. 여기서 예측은 생성형 AI가 산출하는 콘텐츠도 포함함.

(2) AI 법의 적용 범위

- 합의안은 AI 법이 EU 법률의 범위를 벗어난 영역에는 적용되지 않으며, 어떠한 경우에도 국가 안보에 대한 회원국의 권한이나 국가 안보 업무를 위탁받은 기관에 영향을 미치지 않음을 명확히 함.
- 오로지 군사 또는 국방 목적으로만 사용되는 시스템에는 AI 법이 적용되지 않음.
    - 초안은 ‘오로지 군사 목적으로만 개발되거나 사용되는 AI 시스템’을 적용 대상에서 제외하였으나 국방 목적의 AI 시스템에까지 확대된 것으로 보임.
    - 유럽의 시민사회는 AI 시스템이 군사 또는 국방 목적으로만 사용되는 것이 아니라 범용적으로 사용되는 경우가 많고, ‘군사 또는 국방 목적’의 범위가 모호하기 때문에 이러한 예외를 인정해서는 안된다고 주장하였으나 수용되지 않음.
- 연구 및 혁신의 목적으로만 사용되는 AI 시스템, 업무 목적 외로 AI를 사용하는 사람에게는 이 규정이 적용되지 않음.

(3) 금지된 인공지능

- 초안은 위험기반 접근법을 취하여 인공지능의 위험성을 4가지 수준으로 구분하여 규제 수준을 달리 하고 있는데, 이러한 접근법은 합의안까지 유지됨. 즉,
    - ①허용할 수 없는 위험에 해당하는 인공지능을 금지하고, ②고위험은 사람들의 안전이나 기본권에 부정적 영향을 미치는 인공지능으로 이를 제공하거나 사용하려면 엄격한 의무사항을 준수하여야 함. ③ 인간과

상호작용하는 챗봇, 인간의 감정이나 범주를 탐지하는 인공지능, 딥페이크 등 특정 인공지능은 그 사실을 공지하는 투명성 의무가 적용되며, ④ 최소한의 위험에 해당하는 인공지능은 추가적인 법적 의무 없이 기존의 제품안전법률, 개인정보보호법, 차별금지법 등을 준수하면 허용됨.

- 초안에서 규정하고 있는 금지된 인공지능에는 사람의 무의식을 조종하거나 연령·장애 등 취약집단 취약성을 악용하여 피해를 주는 인공지능, 공공기관의 범용 사회신용점수, 법집행기관이 법원의 사전 허가 없이 공공장소에서 원격으로 생체인식을 하는 경우 등이 포함.

- 합의안에서는 특정 인공지능 시스템이 민주주의와 인권에 미치는 위협을 인식하면서 다음과 같은 인공지능을 금지하는데 합의함.
  - 민감한 특성(예: 정치적, 종교적, 철학적 신념, 성적 지향, 인종)을 사용하는 생체 인식 분류 시스템
    - 초안에 없었지만 새롭게 추가됨. 유럽의 정보인권단체 뿐만 아니라, 유럽 개인정보 감독기구의 협의체인 개인정보보호이사회(EDPB) 및 유럽 기구의 개인정보를 감독하는 개인정보보호감독관(EDPS)의 인공지능 법안에 대한 공동 의견서에서도 “공공 기관과 민간 기관 모두에 대해 개인의 생체 인식(예: 얼굴 인식)을 통해, 정치적 또는 성적 지향 또는 유럽연합 헌장 제21조에서 금지하는 기타 차별의 근거 뿐만 아니라, 인종, 성별에 따라 분류하는 AI 시스템”의 금지를 요구한 바 있음.
    - 의회안에서도 “민감하거나 보호되는 속성 또는 특성(의회안은 성별, 성정체성, 인종, 출신 민족, 이주 또는 시민권 상태, 정치적 성향, 성적 지향, 종교, 장애 또는 유럽연합 기본권 헌장 제21조 및 유럽 개인정보보호법(GDPR) 제9조에 따라 차별이 금지되는 기타 모든 사유가 여기에 포함된다고 함)에 따라 자연인을 분류하거나 이러한 속성 또는 특성을 추론하여 분류하는 생체인식 분류 시스템”을 금지된 인공지능에 포함. 의회는 이러한 시스템이 특히 침입적이고 인간의 존엄성을 침해하며 차별의 위험이 크다고 봄.
    - 의회안에 따르면 ‘생체인식 분류(biometric categorisation)’는 "자연인의 생체인식정보 및 생체인식 기반 데이터, 또는 이들 데이터로부터 추론된 데이터를 토대로 그들의 범주를 할당하거나 그들의 특성 또는 속성을 추론하는 것"으로 정의되는데, 여기서 '생체인식 기반 데이터'란 "자연인의 신체적, 생리적 또는 행동 신호와 관련된 특정 기술 처리의 결과로 생성된 데이터를 의미"하는 것으로, 이는 초안에는 없던 정의임. 이 새로운 정의를 추가한 취지는 생체인식정보는 '개인식별'을 전제로 한 생체정보를 의미하기 때문에, '생체인식 기반 데이터'라는 새로운 정의를 통해 '개인식별이 되지

않더라도' 생체정보에 기반하여 개인을 분류하는 시스템까지 포함하고자 하는 것임. 이러한 의회안의 취지가 합의안에서 반영되었는지 확인이 필요함.

- 한국에서도 민감정보인 생체인식정보를 식별이나 인증 목적으로 개인을 식별하기 위한 생체정보로 규정하고 있는데, 개인식별을 하지 않더라도 생체정보를 활용하여 개인을 분류하기 위한 인공지능 시스템을 어떻게 다룰 것인지 고민이 필요함.

- 얼굴 인식 데이터베이스를 생성하기 위해 인터넷이나 CCTV 영상에서 얼굴 이미지를 무차별 수집하는 행위
  - 금지사유 : 대량 감시의 느낌을 가중시키고 사생활 권리를 포함한 기본권의 심각한 침해로 이어질 수 있음을 근거로 함(의회안 서문 26b).
  - 초안에 없었지만 의회안에서 새롭게 추가되었고 합의안에 반영됨. 이는 클리어뷰(ClearView) AI와 같은 관행을 금지하는 것으로 보임. 클리어뷰 AI는 미국의 얼굴인식 업체로 소셜 미디어를 포함하여 인터넷에 공개된 이미지(200억 개 이상으로 알려짐)를 무단으로 수집하고, 얼굴 이미지를 이 데이터베이스와 비교할 수 있는 소프트웨어를 미국의 법집행 기관에 판매하고 있음.
- 직장 및 교육 기관에서의 감정 인식 시스템
  - 초안에는 없었으며 의회안에서 금지된 인공지능으로 포함되었지만, 의회안보다 그 범위가 축소됨. 즉, 의회안에서는 법 집행, 국경 관리, 직장, 교육 기관의 영역에서의 감정 인식 시스템을 금지했으나, 합의안에서는 법 집행 및 국경 관리가 삭제됨. 이는 각 국가 법집행 기관들의 요구가 반영된 것임.
  - 금지사유 : 의회는 이 분야에서 감정인식 시스템을 금지하는 이유를 다음과 같이 설명함. “얼굴 표정, 움직임, 맥박 주파수, 음성과 같은 감정적, 신체적 또는 생리적 특징을 감지하는 것을 목표로 하는 AI 시스템은 그 과학적 기반에 대한 심각한 우려가 있습니다. 감정이나 감정 표현과 그것에 대한 인식은 문화와 상황에 따라, 심지어 한 개인 내에서도 상당히 다릅니다. 이러한 기술의 주요 단점들은 제한된 신뢰성(감정 범주는 신체적 또는 생리적 움직임의 공통 집합을 통해 신뢰성 있게 표현되거나 명확하게 관련되어 있지 않습니다), 특수성의 결여(신체적 또는 생리적 표현이 감정 범주와 완벽하게 일치하지 않음) 및 제한된 일반화 가능성(상황 및 문화의 영향이 충분히 고려되지 않음)에 있습니다. 신뢰성 문제와 그로 인한 남용의 주요 위험은 특히 법 집행, 국경 관리, 직장 및 교육 기관과 관련된 실제 상황에 시스템을 배치할 때 발생할 수 있습니다. 따라서, 이러한

상황에서 개인의 감정 상태를 감지하기 위해 사용되는 AI 시스템을 시장에 출시하거나, 서비스에 투입하거나, 사용하는 것은 금지되어야 합니다"(의회안 서문 26c).

- EDPB-EDPS는 건강 및 연구 목적의 사용을 제외한, 감정인식 인공지능 시스템의 금지를 권고함.

- 사회적 행동이나 개인적 특성을 기반으로 한 사회적 점수 매기기
  - 금지사유 : 일반적인 목적으로 자연인에 대한 사회적 점수를 제공하는 AI 시스템은 데이터가 원래 생성되거나 수집된 맥락과 무관한 사회적 맥락에서 자연인 또는 자연인 집단 전체를 해롭거나 불리한 대우로 이끌거나 사회적 행동의 심각성에 비해 불균형하거나 정당하지 않은 불리한 대우로 이어질 수 있는 등 차별적인 결과를 초래하고 특정 집단을 배제할 수 있으므로, 이는 존엄성과 차별 금지에 대한 권리, 평등과 정의의 가치를 침해한다고 봄(서문 17).
  - 이러한 시스템의 대표적인 사례로 중국의 사회신용 시스템을 들 수 있음. 이러한 시스템은 초안에서부터 금지된 인공지능으로 규정되어 있었으나, 초안의 문구와 의회안의 문구는 차이가 있음. 예를 들어, 초안은 사회적 점수를 부여하는, 공공기관에 의한 인공지능 시스템을 한정하였으나 의회안은 민간의 시스템도 포함하였음. 사회적 점수를 부여하는 인공지능 시스템이라도 구체적으로 어떠한 요건을 충족해야 적용 대상이 되는지에 따라 그 적용 범위가 달라질 수 있기 때문에 합의안의 구체적인 조문을 확인할 필요가 있음.
  - EDPB-EDPS 역시 소셜 미디어와 같은 민간 기업 역시 방대한 개인정보에 기반하여 '사회적 점수 매기기(social scoring)'를 수행하고 있음을 지적하며, 공공부문 뿐만 아니라 모든 형태의 사회적 점수 시스템을 금지해야 한다고 권고함.

- 인간의 자유 의지를 우회하기 위해 인간의 행동을 조작하는 AI 시스템 / 나이, 장애, 사회적 또는 경제적 상황으로 인해 사람들의 취약점을 악용하는 데 사용되는 AI 시스템
  - 금지사유 : 의회안은 이러한 인공지능 시스템을 금지해야 하는 이유를 다음과 같이 설명하고 있음. "인간의 행동을 실질적으로 왜곡할 목적 혹은 그러한 효과를 목적으로 특정 AI 시스템을 시장에 출시하거나 서비스에 투입하거나 사용하는 행위, 그로 인해 신체적 또는 정신적 피해가 발생할 가능성이 있는 행위는 금지되어야 합니다. 이러한 제한에는 AI 시스템의 지원을 받는 신경 기술이 뇌-컴퓨터 인터페이스를 통해 수집된 신경 데이터를 모니터링, 사용하거나 영향을 미치는 데 사용함으로써 자연인의 행동을 그 사람 또는 다른 사람에게 중대한 피해를 야기하거나 야기할 가능성이 있는 방식으로

실질적으로 왜곡하는 경우에, 이 기술이 포함되는 것으로 이해되어야 합니다. 이러한 AI 시스템은 알려졌거나 예측된 성격 특성, 연령, 신체적 또는 정신적 장애, 사회적 또는 경제적 상황으로 인해 개인이 인지할 수 없는 잠재적 구성 요소를 배치하거나 개인 및 특정 집단의 취약점을 악용합니다. 이러한 시스템은 사람의 행동을 실질적으로 왜곡하려는 의도 또는 그러한 효과를 위하여, 그리고 시간이 지남에 따라 누적될 수 있는 피해를 포함하여 해당 사람이나 다른 사람 또는 집단에 중대한 피해를 야기하거나 야기할 가능성이 있는 방식으로 그렇게 합니다.(의회안 서문 16)"

        - 초안에서부터 금지된 인공지능으로 규정되어 있었으나, 초안의 문구와 의회안의 문구는 차이가 있고 구체적인 요건에 따라 금지 범위가 달라질 수 있기 때문에 합의안의 구체적인 조문을 확인할 필요가 있음. 예를 들어, 초안은 '연령, 신체적 또는 정신적 장애로 인한 취약성'으로 한정하였으나 의회안은 '성격 특성이나 사회적 또는 경제적 상황, 연령, 신체적 또는 정신적 능력을 포함하여 개인 또는 집단의 알려지거나 예측된 특성 등 해당하는 개인 또는 집단의 취약점'으로 그 범위를 확대하였음.
    - 개인에 대한 예측 치안(predictive policing) 일부 사례
        - 개인의 특성 혹은 특징(personal traits or characteristics)에 기반한 범죄 수행 가능성에 대한 예측을 포함하여 일부 예측 치안 금지
        - 예측 치안 시스템은 초안에서는 금지된 인공지능으로 규정되지 않고 고위험 인공지능 시스템(초안의 부속서III 6(e))으로 규정되었으나, 의회안에서는 금지된 인공지능으로 규정됨(이런 시스템은 특정인 또는 특정 집단에 대한 차별의 특별한 위험을 가지고 있고, 무죄 추정의 핵심 법적 원칙뿐만 아니라 인간의 존엄성을 침해하기 때문이라고 설명함. 서문 26a). 그런데 합의안은 "some cases of predictive policing for individuals"라고 표현한 것으로 보아 의회안이 부분적으로만 반영된 것으로 보임.
- 공공장소에서의 얼굴인식 등 생체인식 기술을 통한 감시는 가장 큰 쟁점 중의 하나였음.
    - 금지 사유 : 공개적으로 접근 가능한 공간에서 자연인의 '실시간' 원격 생체인식 식별을 위해 AI 시스템을 사용하는 것은 관련 당사자의 권리와 자유를 특히 침해하고, 궁극적으로 인구 대다수의 사생활에 영향을 미치고, 지속적인 감시의 느낌을 불러일으키며, 공공장소에서 생체인식 식별 기술을 배치하는 일방에 통제할 수 없는 권력을 부여하고, 법치의 핵심인 집회의 자유 및 기타 기본권 행사를 간접적으로 억제할 수 있다고 함. 자연인의 원격 생체인식 식별을 위한 AI 시스템의 기술적 부정확성은 편향된 결과를

초래하고 차별적 효과를 수반할 수 있고, 이는 특히 연령, 인종, 성별 또는 장애와 관련이 있다고 함. '실시간'으로 작동하는 이러한 시스템의 사용과 관련하여 영향이 즉각적이고 추가 확인 또는 수정의 기회가 제한적이라는 점은 법 집행 활동과 관련된 사람의 권리와 자유에 미치는 위험을 높인다고 함(서문 18).

- 합의안은 공공장소에서 법 집행 목적으로 생체 인식 시스템을 사용하는 것과 관련하여, 사전에 법원의 허가를 받고 엄격하게 제한된 범죄에 대해서만 시행하는 것을 조건으로 제한적으로 허용함.
- "사후적" 원격생체인식은 심각한 범죄를 저지른 혐의로 유죄 판결을 받았거나 의심되는 사람을 대상으로 하는 표적 검색에만 사용하도록 제한됨.
- "실시간" 원격 생체인식은 엄격한 조건을 준수해야 하고 제한적인 시간과 위치에서 다음과 같은 목적을 위해 사용하는 것으로 제한됨.
  - 피해자(납치, 인신매매, 성 착취)에 대한 표적 검색
  - 구체적이고 현존하는 테러 위협의 예방
  - 법에 언급된 특정 범죄(예: 테러, 인신매매, 성 착취, 살인, 납치, 강간, 무장 강도, 범죄 조직 가담, 환경 범죄) 중 하나를 저지른 것으로 의심되는 사람의 위치 파악 또는 신원 확인
- 초안은 "법 집행 목적으로 공개적으로 접근 가능한 공간에서 '실시간' 원격 생체 인식 시스템의 사용"을 금지했지만, (i) 실종 아동을 포함한 범죄의 잠재적 피해자에 대한 표적 수색, (ii) 자연인의 생명 또는 신체적 안전에 대한 구체적이고 실질적이며 임박한 위협 또는 테러 공격의 방지, (iii) 3년 이상의 최대 기간 동안 구금형 또는 구금 명령으로 처벌 가능한 범죄 행위의 범인 또는 용의자 탐지, 소재 파악, 식별 또는 기소의 경우에는 예외적으로 허용했음. 이에 대해 이러한 예외 허용은 사실상 실시간 원격 생체인식 시스템의 항상적 사용을 허용하는 것이라는 비판이 제기되었음. 또한, 실시간의 요건 자체가 모호하고 사후 생체인식 역시 마찬가지로 인권 침해적임에도 불구하고 이를 금지된 인공지능에 포함하지 않았다는 비판이 제기됨. 이에 의회안에서는 실시간 원격 생체인식 시스템의 사용에 대한 예외 조항을 삭제하고, 사후 원격생체인식도 금지된 인공지능에 포함(다만, 사법적 승인을 받은, 중범죄와 관련된, 표적 수색은 예외적으로 허용)하였음. 합의안은 초안과 의회안의 타협안이라고 할 수 있음. 합의안은 초안에 비해 법집행기관을 위한 예외의 범위를 좁히기는 했지만, 여전히 AI 시스템이 국가의 시민감시의 수단으로 사용될 위험성이 존재함.

(4) 고위험 인공지능

- 초안에서 고위험 인공지능은 기계류, 의료기기, 장난감, 항공, 자동차 등 안전에 미치는 고위험 분야와 원격 생체인식, 교육, 고용, 사회복지·신용평가·응급서비스 등 필수서비스, 경찰, 출입국, 사법 분야에서 사용되는 인공지능으로 기본권에 미치는 고위험 분야가 해당됨.
- 초안에서 고위험 AI 시스템은 '개인의 건강, 안전 및 기본권에 중대한 악영향을 미치는 시스템'에 해당하는 것으로 규정했던 것이 의회안에서는 고위험 AI 시스템은 기본권, 민주주의, 법치 또는 환경을 비롯하여 유럽연합 법률에 의해 인정되고 보호되는 중요한 유럽연합 공익에 허용할 수 없는 위험을 초래하지 않아야 한다고 규정하여(의회안 서문 27), 민주주의와 법치, 환경과 공익에 위험을 초래하는 인공지능을 추가적으로 규정하였음.
    - 초안과 의회안은 AI시스템을 고위험으로 분류, 지정하는 기준으로 AI 시스템이 기본권에 미치는 부정적 영향의 정도를 고려하도록 함. 이러한 권리에는 인간의 존엄성, 사생활 및 가족생활의 존중, 개인정보의 보호, 표현과 정보의 자유, 집회와 결사의 자유, 차별금지, 소비자 보호 교육의 권리, 노동자의 권리, 장애인의 권리, 성평등, 지적 재산권, 효과적인 구제와 공정한 재판을 받을 권리, 방어권과 무죄 추정의 권리, 좋은 행정에 대한 권리, 어린이들이 EU 헌장, 유엔 아동권리협약(디지털 환경에 관한 UNCRC 일반 논평 제25호도 포함)에 명시된 권리. 사람들의 건강과 안전 환경 보호에 대한 기본권을 열거함(의회안 서문 28a).
- 고위험 인공지능 시스템의 의무로는 △위험 관리, △데이터 세트의 품질 관리를 위한 데이터 평가와 데이터 거버넌스, △기술 문서화 및 기록, △인공지능을 사용하는 자에 대한 정보 제공, △인적 감독, △견고성·정확성·사이버 보안의 요구사항을 준수하도록 함. 특히 고위험 인공지능 시스템의 제공자는 출시 전에 △품질 관리 시스템을 구축하고 시판 후 모니터링 시스템을 구축하며, △기술 문서 및 로그 기록을 작성하고, △필요한 적합성 평가 절차를 이행하고 CE 인증을 받아야 하며, △EU 고위험AI 데이터베이스에 등록하고 규제기관에 협력해야 함.
- 고위험 인공지능이 의무사항을 준수하였음을 입증하는 절차는 사전과 사후로 구분됨. 출시 전에는 사전 적합성평가를 이행하되, 안전제품과 원격 생체인식의 경우 제3자가 평가하고 그밖의 고위험 인공지능은 자체적으로 적합성을 평가함. 출시 후 사고가 발생하거나 의무사항을 준수하지 못하는 경우 규제기관이 사후적으로 개입하여 조사하고, 이를 위한 추적가능성과 피해자 권리구제를 보장하기 위하여 문서기록 등 투명성 의무를 강조함.
- 유럽연합 이사회는 "광범위한 고위험 AI 시스템이 승인될 수 있지만, EU 시장에 접근하기 위해서는 일련의 요건과 의무가 적용된다. 이러한 요건은 공동 입법자들이 데이터 품질 또는 중소기업이 고위험 AI 시스템이 요건을 준수한다는 것을 입증하기 위해 작성해야 하는 기술 문서와 관련하여 기술적으로 더 실현 가능하고 이해관계자가 준수하는 데 부담을 덜 주는 방식으로 명확히 하고 조정했다"고

언급함. 이는 AI 법이 인공지능의 개발과 도입을 과도하게 규제하여 혁신을 저해할 것이라는 비판을 의식한 언급으로, AI 법의 규제가 기술적으로도 실현 가능하고 큰 부담이 되지 않을 것이라는 것임.

- 합의안은 건강, 안전, 기본권, 환경, 민주주의, 법치에 잠재적으로 심각한 영향을 미치는 인공지능을 고위험 인공지능으로 분류하면서 준수해야 할 명확한 의무를 부과하는데 합의하였음
    - 이러한 의무에는 의무적인 기본권 영향평가가 포함됨
    - 보험 및 은행 분야, 선거 및 투표 행위에 영향을 미치는 인공지능도 고위험 인공지능에 포함됨. 이는 초안에 없었으나 의회안에서 새롭게 고위험 인공지능에 포함된 것인데 합의안에도 반영되었음.
    - 초안에 없었지만 의회안에서 새롭게 추가된 고위험 인공지능이 더 있는데, (금지된 인공지능에 포함된 것 외에) 감정인식 시스템을 포함하여, 생체인식정보 또는 생체인식 기반 정보를 기반으로 자연인의 개인적 특성을 추론하는데 사용하는 AI 시스템, 시험 중 학생의 금지된 행동을 모니터링하고 감지하는 데 사용하려는 AI 시스템, 자연인을 감지, 인식 또는 식별할 목적으로 국경 관리 활동의 맥락에서 데이터를 모니터링, 감시 또는 처리하기 위해 사용하려는 AI 시스템, 이주 이동 및 국경 통과와 관련된 동향을 예측하거나 예측하기 위해 사용하려는 AI 시스템, [디지털서비스법] 상 초대형 온라인 플랫폼(사용자 4,500만명 이상)으로 지정된 소셜 미디어 플랫폼이 플랫폼에서 사용 가능한 사용자 생성 콘텐츠를 서비스 수신자에게 추천하기 위해 추천 시스템에 사용하려는 AI 시스템 등임. 이러한 의회안이 추가한 고위험 인공지능 목록이 합의안에서 어떻게 처리되었는지 확인이 필요함.
- 법 집행 기관이 긴급한 경우 적합성 평가 절차를 통과하지 않은 고위험 AI 도구를 배포할 수 있는 긴급 절차가 도입되었음.
- 의회안은 고위험 인공지능 시스템에 대한 요구사항으로 에너지 소비량 등을 문서화하고 배출량 등을 투명하게 공개하도록 할 것과 장애인 접근성에 대한 보장 의무를 추가하였는데, 합의안에서 어떻게 반영되었는지 확인이 필요함.
- 고위험 인공지능에 대한 사용자의 의무가 강화되고 인공지능의 영향을 받는 사람에 대한 권리 구제 절차가 신설되었음. (이에 대해서는 아래에서 서술)

(5) 범용 인공지능

- 2022. 11. 챗GPT3 공개 이후, 대규모 언어 모델 혹은 범용AI에 대한 놀라움과 우려가 크게 확산됨.

    - 범용AI에 대한 우려 확산으로 인공지능의 위험성에 대한 국제적 논의가 더욱 활발해짐. 인간의 능력을 초월하는 자율적 초인공지능 등장으로 인류의 실존이 위협받고 있다며 개발 중단에 대한 요구도 등장함.
    - 2023. 10. 30. G7 국가들은 '첨단 AI(Advanced AI)' 회사가 준수해야 하는 자율적 행동 강령에 합의함. 2023. 11. 2. 영국에서도 '프런티어 AI'에 대하여 안전 규정을 마련하려는 AI 안전 정상회의가 개최되어 28개국이 공동으로 '블레츨리 선언'을 채택함.
    - 다른 한편 챗GPT를 비롯한 생성형 인공지능이 제품을 의인화하여 소비자를 위법적으로 기만하였으며, 초인공지능의 먼 미래보다 가까운 문제 해결을 촉구하는 제안도 많아짐. 바이든 행정부 연방거래위원회 (FTC)는 챗GPT의 소비자 기만 여부에 대한 조사에 착수함.
- 챗GPT3 출시 이전에 발표된 초안은 범용 AI에 대한 규정을 포함하고 있지 않았음. 그래서 챗GPT3 공개 이후, 이러한 범용 AI(general purpose AI, GPAI) 및 그 기반이 되는 파운데이션 모델(Foundation model)을 어떻게 규율할 것인지가 인공지능 법안의 첨예한 쟁점이 되었음.
    - 이사회의 공동입장에서는 범용 AI(General Purpose AI, GPAI) 시스템 및 GPAI가 또다른 고위험 시스템에 통합되는 상황에 대응하기 위한 새로운 규정을 포함함. 즉, 고위험 AI 시스템의 일부 요건을 GPAI에 적용하되, 그러한 요건을 직접 적용하기보다는 그러한 시스템에 대한 면밀한 영향평가 및 협의에 기반하여, 그리고 시스템의 특성, 기술적 실현가능성, 시장 및 기술 발전을 고려하여 이행 법률을 제정하는 방안을 제안함.
    - 의회안 역시 '파운데이션 모델(foundation model)'에 대하여 고위험 인공지능과 유사한 의무를 일부 도입함. △위험영향 평가와 완화, △데이터 편향성 평가 및 데이터 거버넌스, △성능과 보안 등 기술 표준 준수, △환경 표준 준수, △정보 공개, △기술 문서 보관을 요구함. △생성형 모델은 콘텐츠가 사람이 아닌 AI 시스템에 의해 생성되었다는 사실에 대한 투명성을 보장해야 함.
- 협상 과정에서 범용 AI에 대한 규제 방안을 둘러싸고 유럽연합 국가간 이견이 컸지만 결국 범용 AI 및 파운데이션 모델에 대한 구체적인 규율에 합의하였음. 합의안에 따르면 파운데이션 모델은 출시되기 전에 특정 투명성 의무를 준수해야 함. 여기에는 기술 문서의 작성, EU 저작권법의 준수, 훈련에 사용된 콘텐츠에 대한 상세한 요약 제공 등이 포함.
- 또한, 영향력이 큰(high-impact) 파운데이션 모델에 대해서는 더 엄격한 규율을 도입함. 이는 대량의 데이터로 학습되고 복잡성, 기능, 성능이 평균을 훨씬 뛰어넘어 가치 사슬을 따라 시스템적 위험을 전파할 수 있는 파운데이션 모델을 의미함.
    - 이러한 모델이 특정 기준을 충족하는 경우 모델 평가를 수행하고, 시스템적 위험을 평가 및 완화하고, 적대적 테스트를 수행하고, 심각한 사고에 대해

집행위원회에 보고하고, 사이버 보안을 보장하고, 에너지 효율성을 보고하도록 함. 또한 이와 관련된 EU 표준이 발표되기 전까지는 이 법을 준수하기 위해 실천 강령에 의존할 수 있음.

(6) 거버넌스

- 초안은 유럽연합 차원의 협력과 조율을 위해 각 국가 감독기관(national supervisory authority)과 유럽개인정보보호감독관(EDPS)으로 구성되는 유럽인공지능위원회(European Artificial Intelligence Board)를 설립하도록 하고 있음. 위원회의 의장은 집행위원회가 맡고 있으며, 위원회는 집행위원회에 대한 자문 역할을 하고 있음. 각 국가는 인공지능 법과 관련된 각 국의 활동을 조율하고 단일 연락소 역할을 할 국가 감독기관을 비롯하여, 여러 관할 기관(national competent authorities)을 지정할 수 있음.
- 이에 반해 유럽의회는 유럽 인공지능 사무소(European Artificial Intelligence Office : 이하 'AI 사무소)의 신설을 제안함. AI 사무소는 유럽연합의 법인격을 갖는 독립적인 기구이며, 운영이사회, 사무처, 자문포럼으로 구성됨. AI 사무소는 회원국, 국가 감독기관, 집행위원회 등에 대한 자문 및 지원 역할과 함께 훨씬 적극적이고 다양한 역할을 수행하는 것으로 규정하고 있음. AI 사무소의 운영이사회는 각 국의 국가감독기관, 집행위원회, EDPS, 사이버보안청과 더불어 유럽기본권청으로 구성되고, 다수결로 의장을 임명함. AI에 관한 규율을 할 수 있는 역량을 대폭 강화하고 권한을 명문화하여, 유럽연합 차원의 독립적인 기구로 AI 사무소의 역할을 강화하고, 집행위원회의 권한을 축소한 것으로 보임.
- 합의안은 AI 사무소 규정을 포함하고 있지만 집행위원회 내에 설립하는 것으로 하고 있고, AI 이사회는 AI 사무소의 운영위원회가 아니라 초안과 같이 집행위원회에 조언을 제공하는 별도의 기구로 설립됨. 즉, 초안에서 유럽의회의 제안을 일부 수용하였지만, 의회안보다는 집행위원회의 주도적인 역할을 인정하는 방향으로 타협이 된 것으로 보임.
    - 합의안은 최첨단 AI 모델을 감독하고 표준 및 테스트 관행을 육성하는 데 기여하며 모든 회원국에서 공통 규칙을 시행하는 임무를 맡은 AI 사무소를 집행위원회 내에 설립함.
    - 독립적인 전문가들로 구성된 과학 패널(scientific panel of independent experts)은 파운데이션 모델의 기능을 평가하는 방법론 개발에 기여하고, 영향력이 큰 파운데이션 모델의 지정 및 출현에 대해 조언하며, 파운데이션 모델과 관련된 물질적 안전 위험을 모니터링하는 등 GPAI 모델에 대해 AI 사무국에 조언을 제공함.
    - AI 이사회(AI board)는 회원국의 대표로 구성되며 집행위원회의 조정 플랫폼이자 자문 기구로서 역할함.

- 업계 대표, 중소기업, 스타트업, 시민사회, 학계 등 이해관계자를 위한 자문 포럼이 설립되어 AI 이사회에 기술적 전문성을 제공함.

(7) AI 시스템 사용자(배치자)의 의무

- 합의안은 고위험 AI 시스템의 제공자 뿐만 아니라, 사용자(혹은 배치자, deployer)의 안전조치 의무도 강화하고 있음. 즉, 배치자는 고위험 AI 시스템을 시장에 출시하기 전에 기본권 영향 평가를 실시해야 함.
  - 초안에서는 '사용자(user)'라는 개념을 사용하였으나, 의회안은 이를 배치자(deployer)로 수정함. 예를 들어, 의료용 인공지능이 있다면 이를 개발한 사업자는 개발자, 이를 실제로 사용하는 병원이나 의사는 배치자(deployer), 환자는 인공지능의 '영향을 받는 사람'이 될 것임. 의회안은 배치자 개념과 함께 '영향을 받는 사람'(affected person)의 개념을 도입하였는데, 합의안에 권리 구제에 대한 내용이 포함된 것으로 보아 합의안에도 반영되었을 것으로 추정됨.
  - 의회안에서는 고위험 AI 시스템의 배치자로 하여금 인적 감독을 실시하고, 감독인이 적절한 자격을 갖추고 훈련을 받도록하며, 사이버보안 조치를 취할 것을 의무화함. 작업장에서 고위험 AI 시스템을 도입할 경우 사전에 노동자 대표와 협의하고 영향을 받는 노동자에게 시스템의 적용대상이 될 것임을 고지해야 함. 또한, 유럽연합 개인정보보호법(GDPR)에 따른 개인정보보호 영향평가를 실시하고 공개할 의무를 부여하고, 나아가 인공지능의 배치로 영향을 받는 사람들의 인권에 미치는 영향을 평가하고 완화하는 기본권 영향평가를 실시하고 공개할 의무를 부여함. 기본권 영향평가를 실시해야 할 배치자의 의무가 합의안에 반영된 것은 매우 고무적임. (인공지능 인권영향평가에 대한 자세한 내용은 2022년 국가인권위원회의 연구보고서 참조)
  - 참고로 고위험 인공지능의 제공자는 출시 전에 적합성 평가 의무가 있음.
- 또한 합의안은 고위험 AI 시스템의 사용과 관련하여 투명성을 강화하는 내용도 담고 있음. 특히, 공공 기관인 고위험 AI 시스템의 특정 사용자도 고위험 AI 시스템의 사용에 대해 EU 데이터베이스에 등록할 의무를 부여하고 있음.
  - 원래 초안에서는 고위험 AI 시스템의 제공자에 대해서만 등록 의무를 부여하고 있었는데, 의회안에서는 고위험 AI 시스템을 사용하는 공공기관과 [디지털시장법] 상 게이트키퍼에 해당하는 배치자에게 등록 의무를 확대함. 그러나 합의안에서는 게이트키퍼 사업자의 등록 의무가 포함되지 않은 것으로 보임.
- 또한 감정 인식 시스템의 사용자(deployer)에게도 자연인이 그러한 시스템에 노출될 경우 이를 알려야 할 의무(투명성 의무)를 부여함.

(8) 권리 구제

- 합의안은 시민들(AI 시스템에 영향을 받는 사람들)이 AI 시스템에 대해 권리구제를 신청할 권리를 명확히 함. 이러한 진정은 관련 시장 감시 기관에 제기할 수 있으며, 제기된 진정은 해당 기관의 고유 절차에 따라 처리됨. 합의안은 자신의 권리에 영향을 미치는 고위험 AI 시스템에 기반한 결정에 대해 설명을 들을 수 있는 권리도 규정함.
- 이는 초안에는 없었지만 의회안에서 새롭게 포함되었고, 합의안에 반영된 것임.
    - 의회안은 “법적 효과를 발생시키거나 이와 유사하게 건강, 안전, 기본권, 사회경제적 복지 또는 본 규정에 규정된 의무에서 파생되는 기타 권리에 부정적인 영향을 미치는 것으로 간주되는 방식으로 자신에게 중대한 영향을 미치는 고위험 AI 시스템의 산출물에 근거하여 배치자(deployer)가 내린 결정에서 영향을 받는 사람은, 제13조 제1항에 따라 배치자에게 의사결정 절차에서 AI 시스템의 역할, 내린 결정의 주요 매개 변수 및 관련 입력 데이터에 대한 명확하고 의미 있는 설명을 요청할 권리가 있다”고 규정하고 있음.
    - 유럽연합 개인정보보호법 제22조(프로파일링을 포함한 자동화된 개별 의사결정)에서도 개인에게 법적 효과를 초래하거나 유사하게 중대한 영향을 미치는 '오로지 자동화된 처리에만 의존하는 결정'에 대해 정보주체가 인간의 개입을 요구하고, 자신의 관점을 표현하고, 결정에 대한 설명을 요구할 권리 등을 보장하고 있음. 이 조항은 인공지능에 의한 자동화된 결정에 대해 그 영향을 받는 사람의 권리를 보장하기 위한 목적이긴 하지만, ‘오로지 자동화된 처리에만 의존하는 결정’으로 엄격하게 규정하고 있음. 이에 비해 의회안에서는 ‘자신에게 중대한 영향을 미치는 고위험 AI 시스템의 산출물에 근거하여 배치자(deployer)가 내린 결정’에 대한 설명 요구권을 폭넓게 인정하고 있음.

(9) 벌칙

- 초안에서는 금지된 인공지능 규정 위반에 대해서는 최대 3천만 유로 또는 전 세계 연간 총 매출액 6% 중 큰 금액의 과징금. 관할당국 협력의무 위반시 최대 2천만 유료 또는 전 세계 연간 총 매출액의 4% 중 큰 금액의 과징금, 부정확한 정보 제공 등의 경우에는 최대 1천만 유로 또는 전 세계 연간 총 매출액의 2% 중 큰 금액의 과징금 부과.

- 합의안에서 AI 법 위반에 대한 과징금은 위반 기업의 직전 회계연도 전 세계 연간 매출액 또는 미리 정해진 금액 중 더 높은 금액의 비율로 책정됨. 금지된 AI 규정을 위반한 경우 3,500만 유로 또는 7%, AI 법의 의무를 위반한 경우 1,500만 유로 또는 3%, 부정확한 정보를 제공한 경우 750만 유로 또는 1.5%가 부과됨. 과징금 액수를 위반 행위의 중대성에 비례하여 조정하였음.

4. 관련 자료

- 인공지능 법안, 집행위원회의 초안
- 유럽연합 이사회(Council of the European Union)의 인공지능 법안에 대한 공동 입장
- 유럽의회의 인공지능 법안에 대한 협상안
- 잠정합의 후 유럽연합 이사회 보도자료
- 잠정합의 후 유럽의회 보도자료
- EDPB-EDPS 공동의견서
- 인공지능 인권영향평가에 대해서는 2022년 국가인권위원회의 연구보고서