한국의 개인정보영향평가와 AI 인권영향평가 현황[1)]

김병욱 변호사(민변 디정위)

1. 들어가며

인공지능의 활용 범위가 넓어지고 그 정도도 심화함에 따라 인공지능 시스템에 내재된 위험이나 인공지능시스템이 적용됨에 따른 파생된 위험 등 부정적 영향에 대한 우려가 커지고 있다.

인공지능 시스템은 대규모 데이터의 학습을 전제하는데, 매개변수의 생성과정은 불투명하고, 자율성을 지니며, 편향성과 오류가 존재할 가능성이 있다. 인공지능 기술의 불투명성, 자율적 특성, 기술의 개발과 활용에 따른 영향, 파급효 등을 고려할 때 사후적인 피해구제나 제재가 어렵다는 특성을 갖는다. 따라서 인공지능 기술이 도입되기 이전 단계에서부터 사전에 위험요인을 식별하고, 위험을 완화, 제거하기 위한 체계적인 시스템이 필요하다.

인공지능 기술이 개인정보보호의 문제 뿐 아니라 광범위한 기본권에 영향을 미친다는 점에서 포괄적인 영향평가 제도로서 인공지능 인권영향평가제도가 시급히 도입될 필요성이 있으며, 아울러 개인정보보호 등 각 분야에서 독자성과 전문성을 발휘하여 인공지능 기술이 가져올 수 있는 부정적인 영향과 위험성을 실효적으로 식별하고, 위험을 통제할 수 있어야 할 것이다.

그러나 현행 국내 인공지능에 대한 인권영향평가제도의 도입수준과 내용, 개인정보영향평가제도의 현황을 살펴보면, 인공지능기술이 새롭게 야기하는 개인정보보호 등 인권에 대한 위협과 부정적인 영향에 대한 실효적 식별, 위험 통제 기능을 제대로 수행하고 있지 못하다.

이하에서 국내 인공지능 인권영향평가 및 개인정보영향평가 제도의 현황에 대해서 검토하여 본다.

---

1) 본 발제문은 유승익 외, 인공지능 인권영향평가 도입방안 연구, 국가인권위원회(2022)의 내용을 일부 포함하고 있다.

2. 국내 인공지능 인권영향평가 현황

가. 인공지능에 대한 인권영향평가 필요성

인공지능 기술에는 기본권 전반에 대하여 부정적 영향을 끼칠 가능성이 잠재되어 있다. 인공지능 시스템은 인권과 기본권 목록의 거의 모든 권리에 침해가능성을 갖는다[2]. 예를 들어 인공지능 알고리즘이 편향된 데이터를 학습하여 차별적인 결정을 내릴 수 있고, 인공지능 시스템이 개인정보룰 무단으로 수집하거나 정보주체의 통제와 무관하게 활용될 위험이 존재한다. 인공지능 기술이 사람의 생명, 신체의 안전, 사생활 보호를 위협할 가능성도 존재한다.

인공지능의 위험성에 대응하여 해외 각국은 다양한 형태의 대응방안을 마련 중인데 그 주요한 장치 중 하나는 인권영향평가를 포함한 영향평가제도이다.

특히 유럽연합은 위험기반 접근법에 따라 인공지능시스템을 수인불가한 위험, 고위험, 제한된 위험으로 차등화하고, 고위험 인공지능 시스템에 대해서는 기본권 영향평가 시행의무를 포함하여 강한 수준의 의무를 부과하고 있다(AI Act 27조). 이외에도 캐나다의 알고리즘영향평가, 네덜란드의 기본권 알고리즘 영향평가, 덴마크 인권영향평가 등 다수의 국가에서 이미 인공지능에 대한 영향평가 또는 인권영향평가를 시행하고 있거나 그 도입을 예정하고 있다.

한편, 유엔 인권기구들은 인공지능 등 신기술이 인권에 미치는 부정적인 영향을 식별·방지·완화하기 위하여 인권실사의 시행을 권고하여 왔고, 인권영향평가가 인권실사의 유용한 도구로서 인권에 미치는 부정적 영향을 식별하고, 대처하는 데 도움이 된다는 점을 지속적으로 강조하여 왔다.

유엔인권최고대표는 인공지능에 의한 많은 추론과 예측은 프라이버시권의 향유에 깊은 영향을 미치고, 사상과 의견의 자유에 대한 권리, 표현의 자유, 공정한 재판 관련 권리 등 다른 권리에도 많은 문제를 야기한다고 하면서, 인공지능시스템의 설계,

2) 유승익, 인공지능이 인권과 민주주의에 미치는 영향과 인공지능법안의 쟁점, 국회토론회 토론집(EU와 미국은 왜 인공지능을 규제하려는가?), 2023. 7.

개발, 배치, 판매, 구입, 운영의 수명주기 전반에 걸쳐 체계적으로 인권실사를 수행하도록 권고하면서, 인권실사의 핵심요소로서 인권영향평가를 제시하였다(2021년 디지털시대 프라이버시권).

국가인권위원회는 2022. 5. 11.「인공지능 개발과 활용에 관한 가이드라인」을 마련하여 국무총리와 관련부처 장관에게 위 가이드라인에 기초하여 인공지능 관련 정책이 수립·이행되고, 관계 법령이 제·개정되도록 할 것을 권고하였는데, 위 가이드라인에서는 인공지능 인권영향평가의 실시를 규정하고 있다.

위 가이드라인은, 인공지능의 개발과 활용에 있어서 인권침해와 차별의 가능성 및 정도, 영향을 받은 당사자의 수, 사용된 데이터의 양 등을 고려하여 인권영향평가를 실시하도록 하면서, 인공지능 인권영향평가 내용에는 인공지능의 특성, 상황, 범위 및 목적을 감안하여 인권 가이드라인이 제시한 원칙 및 내용, 국제 인권 기준, 관련 법률에서 정한 의무 등이 포함되어야 하며, 인권침해 위험요인의 분석 및 개선 사항 등을 등을 도출해야 한다고 정하고 있다.

나아가 인공지능 인권영향평가 결과에서 인권에 미치는 부정적인 영향이나 편향성 및 위험성이 드러난 경우 이를 방지하거나 완화하기 위한 조치사항을 수립하여 적용하여야 하며, 원칙적으로 그 내용이 공개되어야 한다고 정하고 있다.

인공지능의 위험성을 실효적으로 식별하고, 사전에 그 위험 및 부정적인 영향을 완화하기 위하여 인공지능 시스템에 대한 포괄적인 규범체계의 일부로서 인공지능 기술과 인공지능 시스템에 대한 인권영향평가제도를 도입할 필요가 있다.

나. 국내 인공지능 인권영향평가 현황

1) 현행 관련 영향평가제도

영향평가의 안정적인 운영, 절차의 명확성과 의무 부과 등 제도로서의 구속력을 분명히 하기 위해서는 영향평가 시행에 있어 법률적 근거가 필요하다. 그러나 현재 국내에 인공지능 기술이나 서비스만을 독자적인 대상으로 삼고 있는 법제화된 영향평가 제도는 없는 실정이다.

2026. 1.경 시행 예정인 인공지능 발전과 신뢰기반조성 등에 관한 기본법(이하 약칭 ‘인공지능기본법’)에는 고영향 인공지능의 기본권 영향 평가에 대하여 언급하고 있으나, 도입을 강제하고 있지 않고 노력의무를 규정한 수준에 그치고 있다(인공지능기본법 제35조).

다만 관련된 영향평가 제도 가운데 인공지능 기술을 활용한 서비스 또는 사업을 대상으로 일정한 수준에서 영향을 평가할 수 있는 제도가 존재하지 않는 것은 아니다. 다만, 이 경우에도 개인정보자기결정권을 포함한 기본권에 미치는 부정적 영향 또는 위험을 식별하거나, 사전에 예방하는 기능을 수행하고 있거나 수행할 수 있다고 평가하기는 어렵다.

관련하여 주로 거론되는 지능정보화기본법상 사회적 영향평가, 과학기술기본법상 기술영향평가에 대하여 살펴본다.

가) 지능정보서비스 등의 사회적 영향평가

지능정보화기본법 제56조에서는 ‘국민의 생활에 파급력이 큰 지능정보서비스 등의 활용과 확산이 사회·경제·문화 및 국민의 일상생활 등에 미치는 영향’에 대하여 조사, 평가할 수 있다고 하여 ‘사회적 영향평가’를 정하고 있다.

지능정보화기본법상 ‘지능정보서비스 등’에는 전자적 방법으로 학습ㆍ추론ㆍ판단 등을 구현하는 기술로 정의되는 지능정보기술이 포함되어 있고(지능정보화기본법 제2조 제4호 가목), 이는 인공지능 기술을 활용한 서비스를 포함하는 의미로 통용되므로 위 사회적 영향평가는 인공지능 기술을 대상으로 하여 이루어질 수 있다.

지능정보화기본법상 사회적 영향평가의 평가대상에는 ‘안전성 및 신뢰성’, ‘정보보호에 미치는 영향’을 포함하고 있다(지능정보화기본법 제56조).

- 지능정보서비스 등의 안전성 및 신뢰성
- 정보격차 해소, 사생활 보호, 지능정보사회윤리 등 정보문화에 미치는 영향
- 고용·노동, 공정거래, 산업 구조, 이용자 권익 등 사회·경제에 미치는 영향
- 정보보호에 미치는 영향
- 그 밖에 지능정보서비스 등이 사회·경제·문화 및 국민의 일상생활에 미치는 영향

<지능정보화기본법 제56조 제1항 중>

그러나 지능정보서비스 등에 대한 사회적 영향평가는 인공지능 서비스에 대하여 사회적 수용성과 규범적 타당성을 확보하기 위한 목적으로 이루어지는 사후적 평가제도에 가까우며, 인공지능서비스가 사회 전반에 미치는 영향을 매우 광범위한 차원에서 평가한다.

평가 주체도 국가(과학기술정보통신부장관)와 지방자치단체로 한정되어 있고, 평가 범위와 절차에 관하여 시행령 등 관련 법규가 없어 절차가 유동적이며, 평가 결과를 강제할 방법도 없다.

더하여, 평가항목으로 법에 규정되어 있는 '정보보호'는 헌법상 개인정보자기결정권을 충실히 보장하도록 하는 '개인정보보호'와는 목적과 방향성을 달리한다. 지능정보화기본법은 '정보보호'에 대하여 '정보의 수집、가공、저장、검색、송신 또는 수신 중 발생할 수 있는 정보의 훼손、변조、유출 등을 방지하기 위한 관리적、기술적 수단(이하 "정보보호시스템"이라 한다)을 마련하는 것'으로 정의하고 있는데, 이는 '정보보안'과 유사한 개념으로, 정보 주체의 의사를 반영하여 데이터를 통제하고 관리하는 개념의 '개인정보보호'와는 상당히 차이가 있는 개념이다[3].

나) 과학기술기본법상 기술영향평가

지능정보화기본법상 사회적 영향평가제도는 지능정보기술에 대하여 과학기술기본법의 기술영향평가로 대신한다고 정하고 있는데(지능정보화기본법 제56조), 기술영향평가 또한 인공지능이 개인정보보호를 포함한 인권에 미칠 수 있는 부정적 영향을 실효적으로 식별하고 위험을 완화할 수 있는 제도로 평가하기는 어렵다.

3) 서울경제, 2023. 5. 17. " [로터리] 정보보호와 개인정보보호', 'https://www.sedaily.com/NewsView/ 29PM4TS3JV, 검색일 2025. 9. 12.

과학기술기본법 제14조 및 동법 시행령 제23조에 의하면, 기술영향평가의 대상은 ‘미래의 신기술 및 기술적·경제적·사회적 영향과 파급효과 등이 큰 기술로서 과학기술정보통신부장관이 관계 중앙행정기관의 장과 협의하여 정하는 기술’로 평가항목에는 아래 각 사항이 반드시 포함되도록 정하고 있다(과학기술기본법 제14조, 동법 시행령 제23조).

- 해당 기술이 국민생활의 편익증진 및 관련 산업의 발전에 미치는 영향
- 해당 기술이 국민생활의 편익증진 및 관련 산업의 발전에 미치는 영향
- 새로운 과학기술이 경제·사회·문화·윤리 및 환경에 미치는 영향
- 해당 기술이 부작용을 초래할 가능성이 있는 경우 이를 방지할 수 있는 방안
- 해당 기술의 성격과 파급효과가 성별 등 특성에 미치는 영향

<과학기술기본법 시행령 제23조 제2항 중>

기술영향평가는 2003년경 도입된 이래 거의 매년 과학기술정보통신부 주도로 평가가 수행되고 있고, 2015년의 평가대상 기술에 인공지능이 포함되었으며, 2016년은 가상증강현실, 2021년은 레벨4이상의 자율주행자동차였으며 2024년은 ‘안전·신뢰AI기술’이었다.

그러나 기술영향평가는 과학기술정보통신부 주도로 매년 1회 수행되고 있을 뿐, 타 부처에서는 거의 활용되고 있지 못하며, 인공지능을 포함한 신기술의 영향을 인권적 측면에서 평가하기보다는 경제, 사회, 문화, 윤리, 환경 등에 미치는 영향을 전반적으로 평가하여 정책적 제언을 도출하고, 바람직한 정책방향을 유도하기 위한 목적으로 이루어지는 평가라 할 수 있다. 영향평가의 결과를 정책에 반영하도록 규범화되어 있다는 점은 긍정적이나 심도있는 공론 과정없이 형식적인 행정조치에 그칠 가능성이 있다는 점이 지적되기도 한다.

2) 인공지능 자율점검 기준

한편, 법제화된 형태는 아닐지라도, 자율적으로 활용할 수 있는 도구 또는 기준 차원에서 국내에서도 다양한 기준과 도구가 제안되어 왔다.

이들 도구는 시행 여부나 절차에 있어 강제력이나 결과 반영에 대한 구속력은 없다는 한계를 지니고 있으나, 내용적인 면에서 인공지능의 위험성에 대한 문제의식이나

위험성을 식별하는 체계 등의 점에서 참고할만한 부분이 있다고 생각된다.

가) 과학기술정보통신부의 인공지능 윤리영향평가 프레임워크

과학기술정보통신부와 정보통신정책연구원는 2024. 2. 인공지능 윤리영향평가 프레임워크를 발표하였는데, 그 목적을 아래와 같이 제시하고 있다.

- 인공지능 윤리·실뢰성 실천을 위한 기업의 자율적 노력을 지원하고, 사용자가 인공지능을 윤리적이고 주체적으로 활용하기 위한 기준을 제시 - 인공지능 제품·서비스의 윤리적 영향력을 사전에 평가함으로써 긍정적 영향 극대화 및 부정적 영향 최소화를 위한 관리·제도·정책적 조치 방안 등 시사점을 도출 - 인공지능의 윤리적 영향력을 체계적으로 파악할 수 있는 참고자료를 기업, 시민사회, 학계, 공공(정부) 등에 제공하고 인공지능 제품·서비스를 보다 윤리적인 방식으로 개발·배포·활용하도록 장려

인공지능 윤리영향프레임워크에는, 10개 영역별로 구분된 개별 측정문항이 제시되어 있는데, 10개 영역은 과학기술정보통신부와 정보통신정책연구원이 2020년 발표한 「국가 인공지능 윤리기준」의 10대 핵심요건과 유사하게 ① 인권 보장, ② 프라이버시 보호, ③ 다양성 존중, ④ 침해금지 ⑤ 공공성, ⑥ 연대성, ⑦ 데이터 관리, ⑧ 책임성, ⑨ 안전성, ⑩ 투명성으로 명명되어 있다.

그러나 인공지능 윤리영향 프레임워크의 시행 주체는 정부에 한정되어 있으며, 전문성, 공정성, 객관성, 신뢰성, 투명성 확보를 위한 차원에서 산, 관, 학, 연, 평가위원단과 시민사회 등 이해관계자의 참여를 포함시키고 있다는 점 자체는 긍정적이나, 형식적인 의견수렴에 그칠 우려가 있다.

인공지능 기술 또는 서비스의 영향에 대하여 실질적인 평가를 위해서는 해당 서비스의 작동원리, 학습데이터의 출처 등 개발과정의 정보에 대해서도 접근가능해야하나, 대상 인공지능 개발 또는 도입 업체 등이 영향평가에 관여하는지 여부에 대하여 아무런 설명이 없다.

각 영역별 측정문항을 살펴보더라도, 인공지능의 부정적 측면과 긍정적 측면에 대

하여 ‘특정 인공지능 서비스가 인간의 존엄과 가치를 훼손할 위험이 있다’는 수준의 추상적 질의를 하고, 응답자에게 해당 질의에 동의 또는 부동의의 정도를 양적인 척도에 따라 답하도록 하는 방식을 취하고 있어 인공지능의 부정적 영향이나 인권 침해의 위험성을 실효적으로 파악하고 개선할 수 있는 기능은 매우 제한적일 것으로 보인다.

즉, 윤리영향평가에서 ‘윤리’의 의미는 그 자체로 모호하며, 단지 법이나 정책보다 훨씬 느슨한 수준의 적용하거나 준수하기 쉬운 장치의 의미로 사용된 것으로 보인다[4]. 따라서 윤리영향평가 프레임워크를 인권에 대한 영향을 실효적으로 파악하고 위험성을 개선할 수 있는 적절한 도구로 평가할 수는 없다.

나) 국가인권위원회의 인공지능 인권영향평가 도구

국가인권위원회는 2022. 5.경 인공지능 개발과 활용에 관한 가이드라인을 발표한 데 이어, 2024. 5. 인공지능 인권영향평가 도구를 마련하여 과학기술정보통신부장관에게 공공과 민간에 활용되는 모든 인공지능시스템에 대하여 인공지능인권영향평가를 도입할 것을 권고하였다.

현 단계에서 위 인공지능 인권영향평가 도구가 법률로 제도화되지 않아 강제력이 없기는 하나 인공지능 기술의 특성과 인권에 미칠 수 있는 영향에 대하여 다양한 고려사항을 충족시키고 있는 것으로 평가할 수 있다. 국가인권위원회가 제시한 인공지능 인권영향평가도구의 구체적인 내용은 아래와 같다.

(1) 평가대상

인공지능 인권영향평가도구는 공공기관이 도입하는 모든 인공지능시스템과 민간이 도입하는 고위험 인공지능 시스템을 그 대상으로 삼고 있다. 공공분야에서 개발하는 인공지능의 경우 민간 분야에 비하여 그 파급력이 크기 때문에 공공 분야의 경우 모든 인공지능시스템을 평가대상으로 삼되, 민간 분야의 경우에는 범위를 다소 한정하여 고위험 인공지능 시스템으로 대상을 정한 것이다.

---

4) 인공지능 기술의 개발과 보급을 절대선으로 보면서 성문 법이나 규제의 적용을 회피하기 위한 수단으로서 ‘윤리’라는 개념을 사용하고 있는 것이다(허유선 외, ‘왜 윤리인가: 현대인공지능 윤리논의의 조망, 그 특징과 한계’ (2020. 3).

이는 인공지능 시스템의 위험 수준에 따라 요구사항을 차등적으로 적용하는 방식에 따른 것으로, 위험의 완화 내지 제거가 근본적으로 불가능한 '금지되는 인공지능'은 평가대상에서 제외하고 있으며, 고위험 인공지능의 경우에는 영향평가를 실시하도록 정한 것인데, 다만 고위험 인공지능의 구체적인 범위를 명확히 획정하고 있지는 않다.

(2) 평가시기

인공지능 인권영향평가는 시스템이 개발되기 전에 사전 영향평가를 원칙으로 하되, 정기적, 사후적 평가를 통해 지속적으로 위험을 관리하도록 한다.

이는 인공지능의 기반기술이 계속해서 변화하고 있고, 동일한 기술이라 하더라도 운용되는 지정학적, 사회적, 경제적 맥락에 따라 위험성이 달라질 수 있어 사전 영향평가 만으로 인권침해의 가능성이나 위험을 충분히 예방하거나 관리하기 어렵다는 측면을 고려한 것이다.

(3) 평가 주체

인공지능 인권영향평가도구는 개발 주체 및 관련 사업부서와는 독립된 별도의 내부조직 또는 인권 분야 및 인공지능 기술에 대한 전문성과 독립성을 갖춘 제3의 기관에서 수행하는 것이 바람직하다고 밝히고 있는데, 이는 인공지능 시스템의 개발자나 사업의 도입주체에 의한 자체평가의 경우 객관성, 중립성을 담보하기 어렵고, 형식적인 절차로 운영될 가능성이 크기 때문이다.

다만 사업부서와 독립된 별도의 내부조직 또는 인권 분야 및 인공지능 기술에 대한 전문성과 독립성을 갖춘 제3의 기관이 담당하도록 하는 경우에도 인공지능 기술이나 서비스에 대한 정확한 정보나 깊이있는 이해가 결여되어 있을 가능성이 있으므로, 이는 택일의 문제라기 보다는 제도의 전체적인 설계 및 제반 사정을 함께 고려하여 보완하여야할 사항이라고 보인다.

즉, 개발 또는 사업부서가 영향평가를 직접 수행하는 경우 객관성, 중립성을 확보하기 위해 결과 보고서를 공개하거나 사후 점검절차를 두는 형태로 보완할 수 있을 것이다. 또한 제3의 기관이나 독립된 별도의 부서가 영향평가를 수행하는 경우에는 개발 또는 사업 담당부서로부터 원활히 정보를 제공받을 수 있도록 제도를 고안할 필

요가 있을 것이다.

(4) 평가 절차

인공지능인권영향평가는 평가 수행계획을 수립하고 사전 준비하는 '계획 및 준비 단계'(1단계), 부정적 영향에 대한 분석 및 평가를 실시하는 '분석 및 평가 단계'(2단계), 식별된 위험에 대해 방지.완화 및 구제조치를 확인하는 '개선 및 구제 단계'(3단계), 평가에 대해 점검하는 '공개 및 점검 단계'(4단계) 순으로 수행되도록 구성되어 있다.

평가 절차에서 중시되는 점 가운데 하나는 모든 단계에서 이해관계자의 참여를 중시한다는 점이다. 인권영향평가 도구는 이해당사자, 특히 인권이 침해될 가능성이 있는 피해당사자의 참여를 핵심 원칙으로 두면서, 가능한 다양하고 포괄적으로 협의가 이루어져야한다는 점을 강조하고 있다.

이러한 취지에서 각 단계에서 이해당사자에 대한 의견 수렴과 협의 절차를 공통 절차로 포함시키고 있는데, 예를 들어 1단계에서는 이해관계자들과의 협의를 통해 정보를 수집하고 침해가능성이 있는 인권을 식별하는 작업을 수행하도록 정하고 있고, 3단계(개선 및 구제)에서도 인권 침해를 방지, 완화하고 피해자의 권리를 구제하기 위한 의견을 수렴하도록 정하고 있다.

(5) 평가 문항 구성

평가 문항은 평가 절차를 반영하여 4단계로 구성되어 있다.

아래에서 보는 바와 같이 1단계에서 평가대상 인공지능 시스템에 대한 설명 및 이해관계자의 파악을 위한 질의, 인공지능시스템이 도입될 시공간적 맥락, 이해관계자와의 협의관련 자료 확보를 위한 질의를 포함하고 있다.

Q1-2-3. 앞서 파악한, 인공지능 시스템의 이해관계자로부터 해당 시스템이 인권에 미칠 영향에 대한 의견을 수렴하거나 협의하고 이를 문서화 하였습니까.

☐ 예 ☐ 보완 필요 ☐ 아니오 ☐ 정보 없음 ☐ 해당 없음

설명 ( )

Q1-2-4. 이해관계자 의견을 수렴하거나 협의할 때 다음과 같은 내용을 포함합니까.

- 협의한 이해관계자의 성명, 소속, 연락처
- 협의한 일자
- 인공지능 시스템에 대해 이해관계자에게 제공한 자료
- 인공지능 시스템에 대한 이해관계자의 의견

그리고 2단계에서 개인정보보호 및 데이터 관리, 알고리즘의 신뢰성, 차별금지, 설명가능성과 투명성, 자동화 정도와 인간의 개입, 보안, 접근성, 라이선스 등과 관련하여 영향 분석 및 식별에 대한 질의를 하면서, 침해 가능성이 있는 개별 인권을 나열하고, 영향의 범위와 규모 등을 척도로 심각도를 분석하는 내용의 질의를 포함하고 있다.

3단계에서는 인권침해를 방지하거나 완화할 수 있는 조치, 침해를 구제할 수 있는 조치를 검토하고 4단계에서는 영향평가 보고서의 공개 및 점검과 질의로 구성되어 있다.

3) 소결

앞서 살펴본바와 같이 법제화된 영향평가제도 가운데 일부 영향평가제도의 경우 인공지능기술의 영향을 평가할 수 있기는 하나 인권에 미치는 부정적 영향 또는 위험을 실효적으로 식별하고 개선조치를 마련하거나 위험을 통제할 수 있는 기능을 수행하기에는 한계가 있다.

자율점검도구로 제안된 여러 도구들 또한 법적 근거가 없어 강제력을 지니지 못한다는 점에서 한계를 지니나, 국가인권위원회의 인권영향평가도구의 경우 실제로 활용되어 경험을 축적하고, 문제점을 보완하여 나간다면 인공지능 기술, 시스템의 부정적

영향과 위험을 식별, 완화, 방지하는 실질적인 도구로 발전될 가능성은 있다고 볼 수 있다.

3. 국내 개인정보영향평가 현황

가. 인공지능 시스템에 대한 실효적인 개인정보영향평가 필요성

인공지능이 인권에 미치는 부정적 영향 내지 위험성은 다양한 차원에서 논의되고 있지만, 그 중 뚜렷이 가시화된 주요한 위험 중 하나는 개인정보자기결정권에 대한 침해 위험성이라 할 수 있다.

인공지능 기술은 학습데이터, 훈련데이터를 포함한 대량의 데이터를 수집을 전제로 하는데, 이러한 데이터에는 다양한 개인정보가 포함되어 있고, 안면정보나 민감정보를 포함하여 비정형, 정형데이터를 구분하지 않는다. 이를 수집하는 것은 정보주체에 대한 고지 또는 동의 등 합법적인 테두리 내에서 이루어져야 하나 인공지능 개발 과정에서 공개된 정보에 대한 스크래핑 등 무분별한 형태로 과도한 수집이 행해지고 있다.

대량의 데이터에서 개인정보보호를 위해 행하여진 비식별조치는 무의미할 수 있다. 대량의 데이터에서 개인을 식별하거나, 민감한 속성을 추론하는 것이 낮은 비용으로 용이하게 가능할 수 있으며, 생성형 AI를 활용하는 기업은 수집한 정보 뿐만 아니라 온라인에서 확보한 다양한 정보까지 무차별적으로 수집해 새로운 개인정보를 생성하고 있다[5]. 데이터에 기반한 성별, 인종, 연령 등에 대한 추론은 편견, 차별의 위험을 악화시킬 수 있다.

데이터 처리의 불투명성은 정보주체의 권리행사를 불가능하도록 한다. 비정형데이터를 포함한 데규모 데이터셋에서 개인을 식별하는 것이 기술적으로 어렵고 비용이 많이 든다는 이유로 정보주체의 열람, 정정삭제, 처리정지, 파기 등 권리는 실질적으로 보장되지 못하고 있다.

생성형 인공지능의 추론을 통해 생성된 결과에 개인의 속성이나 개인의 정보가 포

5) 장재영, 생성형 AI의 위협과 개인정보자기통제권 보호방안, SW정책연구소(2025. 4)

함되어 있는 경우, 특히 해당 정보가 고용, 승진 등 채용상의 결정, 사회복지 수급자격의 결정과 같은 중요분야에서 활용되는 경우 개인정보 침해 위험을 포함하여 개인의 인권에 미치는 영향이 매우 심대할 수 있다.

따라서 개인정보보호의 측면에서 사전에 인공지능 시스템이 가져올 수 있는 부정적인 영향과 위험성을 실효적으로 식별하고, 위험을 통제할 수 있도록 개인정보 영역의 별도의 영향평가 제도가 포괄적인 규제체제의 일부로 함께 구축될 필요가 있다.

인공지능 시스템의 개인정보보호와 관련한 부정적 영향이나 위험에 대한 평가 항목이나 평가 지표는 인권영향평가와 중복되는 경우가 발생할 수 있으나, 상호 완전히 수렴할 수 없으며 고유의 독자적 영역이 존재한다. 개인정보영향평가가 인권영향평가 제도와 상호 보완적인 형태로 운영된다면 평가 대상인 업체나 개인, 기관에 대하여 중복 규제부담을 지운다고 볼 수도 없을 것이다.

유럽연합은 GDPR상의 데이터보호영향평가의 내용과 동일, 유사한 평가를 실시하였을 경우, 기본권 영향평가는 이를 보완하는 방식으로 이루어지도록 정하고 있고, 덴마크 인권영향평가 또한 개인정보영향평가와 상호 완전히 수렴하지 않는다는 점을 지적하고 있다.

나. 국내 개인정보영향평가 현황 및 문제점

개인정보영향평가 제도는 개인정보보호법 제정과 함께 2011년부터 도입되어 공공부문에서 일정 규모 이상의 개인정보 파일 운용시에 사전 영향평가를 하는 형태로 시행되고 있으나, 실질적으로 개인정보보호에 위협이 되는 위험 요인을 식별하고 개선조치를 도출하는 사전 예방적인 기능을 수행하고 있지 못하다.

인공지능이 제기하는 새로운 위협에 대응하여 개인정보영향평가제도가 실질화될 필요가 있다. 아래에서 현행 개인정보영향평가제도의 문제점에 대하여 보다 구체적으로 검토하여 본다.

1) 의무실시 대상(공공기관) 및 양적 기준의 문제

현행 개인정보영향평가제도는 의무 실시 대상을 공공기관으로 한정하고 있고, 공공기관 외에 민간업체 등 개인정보처리자에 대해서는 시행 여부를 자율에 맡기고 있다(개인정보보호법 제33조 제11항). 공공기관에서 인공지능을 활용하여 개인정보를 처리하는 경우 파급효가 더 크고, 공적 영역에 요구되는 책임성 등을 고려하면 공공기관에 더 무거운 의무를 부과한다는 점은 일응 수긍이 가는 점이 있다.

그러나 의무 실시 대상의 범위는 무엇보다도 개인정보의 침해의 위험성을 기준으로 삼아야 할 것인데, 공공기관이 아닌 민간 영역의 경우에도 고위험 인공지능 시스템을 개발하거나, 이를 도입하거나 활용하는 경우 개인정보보호의 측면에서 심각한 위험이 존재한다. 인공지능 시스템의 개발 및 활용이 주로 민간 영역에서 이루어진다는 점을 고려하면, 민간 업체로 하여금 사전에 위험요인을 식별하고, 위험요인을 완화하거나 제거할 수 있도록 영향평가를 의무적으로 실시하도록 할 필요성이 있다는 점에 이견이 있기 어려울 것이다.

또한 개인정보보호법은 정보주체의 숫자라는 양적인 기준에 따라 개인정보파일을 구축·운용 또는 변경하려는 경우(5만명 이상의 민감정보 또는 고유식별정보의 처리가 수반되는 경우), 개인정보파일과의 연계결과 50만명 이상의 개인정보가 포함되는 개인정보파일의 경우, 영향평가를 마친 후 개인정보 검색체계 등 개인정보파일의 운용체계를 변경하려는 경우 개인정보영향평가의 의무 실시 대상으로 정하고 있는데(개인정보보호법 시행령 제35조), 인공지능시스템의 경우 처리하는 개인정보의 숫자가 적은 경우에도 시스템의 적용분야, 처리 대상 정보의 종류, 작동방식에 따라 부정적 영향의 정도나 심각성이 클 수 있다.

따라서 인공지능 시스템의 경우 위험기반 접근법에 따라 도입되는 분야(예를 들어 공공 장소에서 실시간 생체정보 수집을 통한 신원확인, 채용, 승진 등 인사상 결정에 사용되는 경우 등)에 따라 개인정보영향평가의 의무실시대상에 포함시킬 필요가 있다.

2) 영향평가시 고려할 사항, 평가기준, 평가 항목의 문제

개인정보보호법은 개인정보영향평가시 고려할 사항으로 처리하는 개인정보의 수, 개인정보의 제3자 제공여부, 정보주체의 권리를 해할 가능성 및 그 위험의 정도, 민감정보 또는 고유식별정보의 처리여부, 개인정보 보유기간을 두고 있다(개인정보보호

법 제33조 제3항 및 동법 시행령 37조).

또한 평가기준으로 다음 4가지를 제시하고 있다(개인정보보호법 시행령 제38조 제1항).

- 해당 개인정보파일에 포함되는 개인정보의 종류·성질, 정보주체의 수 및 그에 따른 개인정보 침해의 가능성
- 법 제23조제2항, 제24조제3항, 제24조의2제2항, 제25조제6항(제25조의2제4항에 따라 준용되는 경우를 포함한다) 및 제29조에 따른 안전성 확보 조치의 수준 및 이에 따른 개인정보 침해의 가능성
- 개인정보 침해의 위험요인별 조치 여부
- 그 밖에 법 및 이 영에 따라 필요한 조치 또는 의무 위반 요소에 관한 사항

<개인정보보호법 시행령 제38조 제1항 중>

개인정보영향평가에 관한 고시에서는 영향평가기준에 따른 평가영역과 평가분야를 구체화하고 있는데, 종래 영향평가 기준(평가영역, 평가분야)에서는 인공지능 시스템의 개발이나 활용과정에서 새롭게 생겨나거나 발생할 가능성이 있는 부정적 영향이나 위험성을 전혀 고려하지 않고 있었다.

최근 2025. 9. 5.부터 시행된 고시에 의하면 평가분야에 '인공지능'이 포함되고, 그 세부분야로 '인공지능 시스템 학습 및 개발', '인공지능 시스템 운영 및 관리'로 구분하여 아래와 같이 안내서에서 평가항목을 좀더 구체화하고 있기는 하나[6], 그 자체로 충분하다고 보기는 어렵다.

평가분야(고시)	세부분야(고시)	주요 평가항목(안내서)
인공지능(AI)	인공지능(AI) 시스템 학습 및 개발	■ 개인정보 처리의 적법성 보장 ■ 불필요한 민감정보 등 수집 방지 ■ 학습데이터 보유기간 명확화 ■ AI 취약점 공격에 의한 개인정보 유·노출 등 최소화
	인공지능(AI) 시스템 운영 및 관리	■ AI 개발·운영 주체 간 책임 명확화 ■ 개인정보 처리의 투명성 확보 ■ 생성형 AI 서비스의 허용되는 이용방침 제공 ■ 정보주체 권리보장 방안 수립·시행

<출처 : 보안뉴스 2025. 9. 4.자 기사>

6) 보안뉴스, 2025. 9. 4., '공공기관 AI활용시 개인정보보호 강화된다' 'https://m.boannews.com/<html/detail.html?idx=139079>, 검색일 2025. 9. 12.

위 추가된 항목에 의하더라도 인공지능 시스템이 적용되는 맥락 또는 적용분야에 따라 생겨나거나 증대될 수 있는 개인정보 침해 위험을 고려하지 못하고, 인공지능 시스템에 요구되는 설명가능성, 투명성 등을 충분히 고려하고 있지 못하며, 정보주체의 권리 보장에 대해서도 방안 수립, 시행 여부를 묻는 수준의 형식적인 차원에 그치는 등 미비점이 존재한다.

또한 인공지능 시스템에 고유하게 존재하는 개인정보침해 위험에 대한 안전조치나 구제조치 등도 평가항목에 포함시켜야할 것이나, 이 또한 누락되어 있다는 점에서 한계가 있다.

3) 이해관계자(영향을 받는자)의 절차 참여 부재

인공지능 위험성에 대한 대응 체계 뿐만 아니라 영향평가의 실질화에서 중시되는 사항 가운데 하나는 이해관계자의 참여이다. 국가인권위원회에서 보급한 인공지능 인권영향평가도구에서는 전체 평가절차에서 공통적으로 이해관계자의 참여를 핵심원칙으로 두고 있는데 이는 영향평가의 객관성, 공정성을 도모하고 일방 이해당사자에 치우친 결과를 도출하지 않기 위한 취지이다.

그러나 현행 개인정보영향평가는 인공지능시스템에 의해 개인정보측면에서 영향을 받는 이해당사자인 정보주체의 참여나 협의에 관하여 전혀 고려하지 않고, 이와 관련하여 아무런 규정을 두지 않고 있다는 점에서 한계가 있다.

4) 영향평가 결과보고서의 공개 및 검증 관련

영향평가의 신뢰성 담보를 위해 공개와 검증 절차가 필수적이다. 개인정보보호법은 공공기관이 영향평가를 시행하는 경우, 그 결과를 개인정보보호위원회에 제출한다고 규정하고 있고(개인정보보호법 제33조 제1항), 개인정보파일을 등록할 때 영향평가결과를 함께 첨부하도록 하고 있으며(개인정보보호법 제33조 제5항), 2023. 9. 15. 법개정에 따라 개인정보영향평가의 요약본을 공개할 수 있도록 하고 있다(개인정보보호법 시행령 제38조 및 고시 제12조의 2).

그러나 민간 영역에서 영향평가가 시행되는 경우에 대해서는 영향평가의 공개나 검

증과 관련하여 아무런 규정을 두지 않고 있다.

앞서 언급한바와 같이 민간영역에서 인공지능 시스템이 활용되는 경우 적어도 고위험 인공지능시스템의 경우에는 개인정보영향평가의 수행을 의무화할 필요가 있고, 이 경우 영향평가의 결과를 개인정보보호위원회에 제출하도록 하거나, 요약본을 공개하도록 하는 등의 의무를 부과할 필요가 있을 것이다.

5) 영향평가 시기의 경우

인공지능 기술은 그 기반기술이 계속 변화하고 있을 뿐 아니라, 동일한 기술이라하더라도 운용되는 지정학적, 사회적, 경제적 맥락에 따라 위험성이 달라질 수 있어 인공지능 시스템이 사업에 도입되는 단계 뿐만 아니라 이후에도 시스템의 중대한 변경이 있거나, 시스템의 적용 범위 등에 중대한 변경이 있는 등의 경우에도 지속적인 위험 요인 분석과 관리가 필요하다.

그러나 현행 개인정보보호법에 의하면, 개인정보의 처리를 수반하는 인공지능 시스템을 구축, 운용하기 이전 시점 및 기존 인공지능 시스템을 변경하는 경우 평가를 수행하도록 정하고 있는 것이 전부여서(개인정보보호법 제33조), 지속적인 위험 관리가 불가능한 구조이다.

6) 결과 반영의 의무화 정도가 미약함

영향평가제도가 실질화되기 위해서는 결과보고서 작성에 그치는 것이 아니라, 결과를 반영한 개선조치의 이행을 담보할 수 있는 장치가 필요하다. 영향평가를 통하여, 인공지능 시스템의 부정적 영향이나 위험성을 식별하고, 이를 완화하거나 제거할 수 있는 조치를 도출해내면 이를 인공지능시스템의 설계 및 개발, 활용 과정에 반영하여야 하는 것이다.

그런데 현행 개인정보보호법은 개인정보보호위원회로 하여금 영향평가 결과에 대하여 의견을 제시할 수 있다고 하여 의견 제시여부를 재량사항으로 두고 있고(개인정보보호법 제33조 제4항). 의견을 제시하더라도 그 이행을 담보할 수 있는 수단이 없다.

개인정보보호위원회는 최근인 2024. 11.경 고시를 개정하여 영향평가 결과 개선사

항으로 지적된 부분에 대한 이행결과 및 이행계획 제출 시한을 기존 1년에서 2개월로 단축시키긴 하였으나, 이 역시도 이행을 담보하기에는 충분하지 못하다. 영향평가에서 도출된 개선조치의 반영의무를 구체화된 법규정으로 명확히 하고, 불이행에 대한 제재규정 또한 도입할 필요가 있다.

다. 소결

인공지능기술은 개인정보보호 측면을 포함하여 기본권에 중대한 위협이 되고 있으나, 현행 개인정보영향평가제도는 개인정보의 처리를 수반하는 인공지능 시스템의 광범위한 활용에도 불구하고 도입된 취지에 부합하는 위험 예방 및 통제 기능을 거의 수행하지 못하고 있다.

인공지능이 새롭게 제기하는 위협에 대응하여, 부정적 영향 및 위험의 사전 예방기능을 수행할 수 있도록 개인정보영향평가 제도 전반에 걸쳐 전면적인 개선이 필요하다.

4. 결론

인공지능의 위험성과 부정적 영향을 관리, 통제하기 위해서는 체계적인 시스템이 필수적이다. 개인정보영향평가 제도가 실질화되어, 인공지능 인권영향평가와 함께 인공지능의 위험을 통제하고, 관리할 수 있는 포괄적인 규제 체계의 일부로 실효적 기능을 수행하여야 한다.

<참고 문헌>

유승익 외, 인공지능 인권영향평가 도입방안 연구, 국가인권위원회, 2022
유승익, 인공지능이 인권과 민주주의에 미치는 영향과 인공지능법안의 쟁점, 국회토론회 토론집(EU와 미국은 왜 인공지능을 규제하려는가?), 2023. 7.
이호중 외, 유럽연합「개인정보보호 규정」(GDPR) 등 국제인권기준에 따른 개인정보보호 법제도 개선방안연구, 2020
허유선 외, 왜 윤리인가: 현대 인공지능 윤리 논의의 조망, 그 특징과 한계, 2020. 3.

장재영, 생성형 AI의 위협과 개인정보자기통제권 보호방안, SW정책연구소, 2025. 4
인공지능영향평가도구해설서, 국가인권위원회, 2024. 12.
인공지능 윤리영향평가프레임워크(2024), 과학기술정보통신부·정보통신정책연구원, 2024. 2.