대한의료법학회
『의료법학』 제24권 제4호
https://doi.org/10.29291/kslm.2023.24.4.067

생성형 AI의 의료적 활용과 개인정보보호

이 수 경*

Ⅰ. 들어가는 말
Ⅱ. 생성형 AI 시대의 발생 가능한 쟁점
1. 생성형 AI의 특징
2. 발생 가능한 법적 쟁점
Ⅲ. 생성형 AI의 활용과 개인정보보호
1. 개인정보의 개념과 보호범위
2. 개인정보 보호와 보건의료데이터 관련 법제
3. 개인정보 관련 법제
4. 생성형 AI의 활용과정에서의 개인정보 침해에 대한 제언
Ⅳ. 맺는 말

I. 들어가는 말

올해 가장 화두가 되었던 생성형 AI는 앞으로의 우리에게 많은 변화를 가져올 것으로 보인다. 생성형 AI의 대표적인 예인 GPT는 인공지능 연구회사인 Open AI社의 모델로서 인공지능이 모든 인류에게 이익이 되도록 연구를 하는 것을 기업의 목표로 소개하고 있다. 대화 방식으로 작동하는 모델인 GPT를 통하여 후속 질문에 답변하면서 실수를 인정하고, 잘못된 전제에 대하여 도전하며, 부적절한 요청을 거부하기도 한다.[1] 대화형 인공지능의 생성형 AI기술은

* 논문접수: 2023. 12. 17. * 심사개시: 2023. 12. 18. * 게재확정: 2023. 12. 27.
* 법학박사, 고려대학교 연구교수.
* 이 글은 대한의료법학회 10월 월례학술대회에서 발제한 것으로 플로어에서 주신 여러 귀한 의견을 반영하여 수정하고 보완한 것입니다. 지면을 빌려 감사의 인사를 올립니다. 본 연구는 『2023년 생명윤리정책연구 자유공모 지원사업』에 선정되어 (재)국가생명윤리정책원으로부터 연구비를 지원받았습니다.

의료산업에도 침투에 대하여 의사 전용 지식공유 플랫폼인 '인터엠디(Inter MD)'의 의사 회원 약 천여 명을 대상으로 한 설문조사에서 조사결과 전체 응답자의 43.8%는 의료현장에서의 GPT 활용에 대하여 진단과 처방의 "보조수단"으로만 활용해야 한다고 했으며, 19.2%의 응답자는 신뢰할 만한 의료정보를 학습시켜야 한다고 답하였고, 의학적 판단이 아닌 정보의 단순 정리에 대하여 활용해야만 할 것이라는 응답도 14.8%나 나타났다. 의료현장에서의 GPT의 활용을 반기면서도 또 한편 서류작성이나 반복 업무를 대신하는 것을 기대했다는 응답을 보면 의료진의 단순 노동의 피로도를 줄여주는 데 대한 기대감으로 해석된다.[2] 사실 GPT가 미국의 의사시험을 통과했다는 언론 보도 이후, 과연 의료영역에서도 실제로 활용이 가능할 것인지 검증을 위한 작업이 본격화 되는 모습이다. 비단 의료분야 뿐만이 아니라, 다양한 분야에서 긍정적인 방향으로 활용될 수 있는 가능성[3]도 있지만, 오류가 있는 답변이 제시된 사례도 지적되어 검증에 대한 목소리가 높아지고 있다. 의학논문 검색 사이트인 PubMed에도 GPT 관련 연구가 2022년 12월 기준 141여 건이 등록되었는데, 이는 의학계의 관심도가 상당함을 반증하는 것이다. 또한 초기의 연구는 대체로 의학적 활용에 대한 것이었으나, 이후 임상 현장을 구현하고 GPT의 답변을 검증하는 단계로 접어들고 있다고 평가된다.[4]

인공지능 제조물 혹은 더 정확하게는 인공지능 알고리즘은 그 발달단계에

---

1) OpenAi의 소개에 대하여, OpenAI is an AI research and deployment company. Our mission is to ensure that artificial general intelligence benefits all of humanity. https://openai.com/about; GPT에 대하여, We've trained a model called ChatGPT which interacts in a conversational way. The dialogue format makes it possible for ChatGPT to answer followup questions, admit its mistakes, challenge incorrect premises, and reject inappropriate requests. https://openai.com/blog/chatgpt; 마지막 방문일 2023. 6. 23.

2) 선모은 기자, 이코노미스트, "의료진도 찾는 인공지능?… 챗 GPT는 보조 수단일 뿐", 기사 입력 2023. 5. 9. 마지막 방문일 2023. 10. 1. ※ 이 글에서 마지막 방문일이 별도로 표기되지 않은 온라인 자료는 모두 2023. 10. 1.임을 밝힌다.

3) 생명공학의 발전과정에 인공지능이 결합될 경우 전례 없는 새로운 잠재적인 해결책이 제공될 것이라는 기대로서, Andreas Holzinger, Katharina Keiblinger, Petr Holub, et al.(2023). AI for life: Trends in artificial intelligence for biotechnology. New Biotechnology. 74: 16-24.

4) 최선 기자, Medical Times, "의료영역 챗GPT찾아라...의학계 검증 착수", 기사 입력, 2023. 3. 22. <https://www.medicaltimes.com/Main/News/NewsView.html?ID=1152708>.

따라서 민사책임의 정도가 달라질 수 있다. 단순히 위험하거나 복잡하고 시간이 오래 걸리는 작업 등을 인간을 대신해서 수행하는 산업용 로봇의 경우도 있으나[5], 의료용 로봇의 경우 보다 까다롭고 정교한 작업을 수행해야 할 수 있다. 현재의 의료용 로봇은 단순히 의료진을 물리적으로 보조하는 단계에 머무른다 하더라도 생성형 AI의 시대에는 여러 데이터를 통합하여 수술로봇이 자동화될 수 있을 것이다.[6] 또한 미국 등 주요국에서도 이와 같이 생성형 인공지능의 의료계에서의 활용이 멀지 않은 미래로서 상정되고 있다.[7]

총론적인 관점에서의 생성형 AI에 대한 문제점 중의 하나는 생성형 AI가 근본적으로 대규모 언어모형 인공지능 이라는 점에서 파생될 수 있는 한계점이다. 기본적으로 대규모 언어모형(large-scale language model, 이하 'LLM') 인공지능은 법적인 관점에서 봤을 때 저작권 침해[8]나 데이터베이스 제작자의 권리가 우선적으로 논의 될 수 있는 문제점이지만 본고에서는 대규모 인공지능의 생래적인 특징인 "인공신경망 내부의 학습데이터를 암기(memorization)하고 있다가 그것을 그대로 재현해 내는" 것에 대한 민사법적인 문제점을 지적하고자 한다. 의료 데이터라는 것은 특성상 불가피하게 환자 개인의 특성이 반영되게 마련이고 개인정보를 재생성하는 등의 문제까지도 발생할 가능성이 있다. 기존의 인공지능에서보다 훨씬 방대한 활용이 우려되기에 이러한 문제에 대한 법적인 문제의 발생가능성을 결코 묵과할 수 없는 상황이다.[9] 주어

---

5) 이수경, "산업용 AI로봇의 제조물 책임법 적용을 위한 소고", 정보법학(제26권 1호), 2022, 251면.

6) 송미옥·조용진, "의료로봇의 현재와 미래: 수술로봇을 중심으로", 디지털 융복합연구(제19권 4호), 2021, 349면 이하.

7) L. Eliot. (2023), Generative AI is stoking medicla malpractice concerns for medical doctors in these unexpected ways, says AI ethics and AI law. Fobes. <https://www.forbes.com/sites/lanceeliot/2023/05/23/generative-ai-is-stoking-medical-malpractice-concerns-for-medical-doctors-in-these-unexpected-ways-says-ai-ethics-and-ai-law/?sh=6307f1c96437>.

8) 본고에서 저작권 침해에 대하여 따로 논의하지 않는 이유는, 데이터를 수집하는 과정과 기계학습의 과정에서의 저작권법에 대한 문제는 이제 수용을 할 것인지에 대한 경계선에 있다고 보는 견해 또한 상당하기 때문이다. 저작권에 대해서는 특별히 생성형AI만의 문제로 보기에도 무리가 있다. 이에 대하여 동지의 견해로, 김윤명, "생성형 인공지능(AI) 모델의 법률문제", 정보법학(제27권 1호), 2023., 80면 이하.

진 자연어 이해 과제를 해결하기 위하여 언어모형을 이용하게 되고, 대화형 '챗봇(Chat bot)' 등에서 자동완성 기능을 적용하여 완성된 문장을 생성해 내게 된다. 언어 모형 중에서도 구글의 BERT나 Open AI의 GPT-3은 폭넓게 활용되는 현실이다. 기본적으로는 인간이 입력한 문장에 대한 답변을 도출하는 것으로 기능을 수행하나, 인간이 직접 작성한 것과 거의 동일한 수준으로까지 올라선 것으로 보이고 심지어 소수의 작업 사례만을 제시하더라도 인공신경망이 이를 이해하고 요청된 작업을 수행하기에 이르게 되는데 인공지능이 학습 데이터를 수집하는 방식에서 공개된 개인정보의 범위[10]에 대한 문제점은 지속적으로 문제가 될 수 있다.[11] 인공지능의 생래적인 특징에서 비롯되는 예측불가능성과 인공지능이 빅 데이터를 통하여 학습하는 과정에서 무단으로 웹 페이지의 내용을 그대로 가져와서 크롤링하는 경우 불법행위의 문제가 발생할 여지도 있다. 의료 분야에서 데이터를 가져오는 과정에서 기본적으로 의료 관련 저널이나 학술자료의 저작권 침해에 대한 우려부터 환자 개인의 데이터를 침해할 우려 등은 실정법을 위반할 가능성을 내포하기도 한다.[12] 본 연구는 수집된 데이터를 통하여 인공지능의 시스템을 이용한 행위에 대한 책임에 대한 논의에 앞서 앞으로 의료분야에서 생성형 AI가 활용될 경우의 개인정보 보호와 관련한 논의를 하고자 한다. 개인정보 활용의 문제는 생성형 AI의 의료계의 활용에 있어서 가장 직접적인 문제가 발생할 것으로 생각되는 부분이며 앞으로 민사책임에 대한 논의로의 확장을 위해서도 이 부분에 대한 연구가 선행되어야 할 것으로 여겨진다.

---

9) 대규모 언어모형 인공지능의 법적인 쟁점에 대하여는 김병필, "대규모 언어모형 인공지능의 법적 쟁점", 정보법학(제26권 1호), 한국정보법학회, 2022., 174-176면 참조.

10) 우리나라의 개인정보의 활용에 대하여 그 요건이 지나치게 높고 정보자유와 정보보호사이의 이익형량을 그르쳤다는 견해도 있다. GDPR의 경우와 같이 개인정보보호법 제15조 제1항 제6호를 개정하여 정보주체의 이익이 우월하지 않는 이상에는 처리를 허용하고 다만 제37조를 개정하여 이에 대응하는 처리 정지사유를 추가하는 것이 바람직하다는 견해로, 이동진, "일반적으로 접근 가능한 개인정보의 처리와 이익형량", 정보법학(제24권 2호), 2020., 77면.

11) 앞의 논문, 179면, 186면, 198면.

12) 차상육, "인공지능(AI)과 지적재산권의 새로운 쟁점-저작권법을 중심으로-", 법조(제723권), 2017., 228면 이하.

II. 생성형 AI 시대의 발생 가능한 쟁점

1. 생성형 AI의 특징

생성형 AI는 단시간에 방대한 데이터를 수집하고 분석하므로 수 만개의 임상 케이스를 분석하고 반영해야만 하는 의료분야에서는 활용 가능성이 높다. 그러나 이로 인하여 손해가 발생하였을 경우 과연 그 책임은 어떻게 분배를 하는 것이 공평과 정의의 법원칙에 부합할 것인가. 보통의 민법의 사안들은 기본적으로 판례군이 형성되어 있어서 법관에 의한 법형성(richterliche Rechtsforbildung)의 과정을 추적하여 연구하거나, 또는 법적 삼단논법을 연구하여 최종적으로 해석 작업을 통하여 획득된 법적인 명제에 대하여 사안을 적용해 보는 형식을 취하는 것으로 연구의 방법을 살펴볼 수 있다.[13] 그러나 생성형 AI의 경우에는 기존에 법관에 의한 법 형성을 논의할 만한 사례군이 집적되어 있지 않고, 더구나 법명제로 분석할 만한 특별법제도 만들어져 있지 않은 경우가 대부분이다. 따라서 이 경우, 발생 가능한 손해를 가상으로 설정해 볼 수 있을 것이다.

GPT는 일반적으로 다음의 특징을 가지는 것으로 파악된다. 인공지능 연구회사인 Open AI에 의하여 만들어진 GPT는 2022년 11월 출시되었고, Open AI는 여러 투자자에 의해 지원을 받고 있다. 특수한 알고리즘을 사용하여 데이터 시퀀스 내에서 패턴을 찾는 'Generative Pre-trained Transformer'를 통한 작동을 의미한다. 알고리즘에 대한 미세 조정 프로세스가 포함된 GPT-3.5모델이 현재 사용되고 있으며, 더 빠른 응답시간과 인터넷 플러그인을 제공되는 GPT-4는 사진에 대한 설명, 이미지 캡션의 생성, 최대 2만 5천 단어의 응답 생성 등이 처리되어 더욱 복잡한 작업을 처리할 수 있게 된다. 기본적으로 딥러닝을 활용한 변환기 신경망(transformer neural networks)은 인간과 유사한 텍스트를 생성하고 이러한 신경망의 일반적인 시퀀스를 기반으로 다음 단어, 문장, 텍스트 등을 예측한다. GPT는 인간의 언어를 배우기 위하여 온라인

13) 법 해석론에 대한 기준으로서는 이계일, "독일의 법해석론에 대한 연구", 연세법학(제38권), 연세법학회, 2021., 655면 이하.

텍스트로 훈련을 받고 대화의 기본을 배우기 위해서는 대본을 사용했다. 결국 이 프로그램은 자연어 처리를 사용하여 인간과 같은 대화를 생성하는 인공지능 챗 봇으로 언어 모델로서 질문에 응답하고 소셜 미디어나 게시물, 에세이, 코드 및 이메일을 포함한 다양한 형태의 서면 콘텐츠를 구성할 수 있다. 이용자가 질문을 하거나 답변에 대한 설명을 요청할 수 있는 점에서는 고객 서비스의 웹사이트 자동채팅 서비스와 유사하나, 인간의 피드백과 최고의 응답순위를 선정하는 보상모델을 통한 강화학습으로 훈련되고 응답을 개선하는 과정을 거친다.[14] GPT-3 의 경우, 기본적 기능이 인간과 유사한 수준으로 작문을 수행하는 것으로서 기술적으로는 대규모 언어모형(large-scale language model)이다. 문장이 해당 언어에서 존재할 확률을 부여하는 모형을 '확률모형'이라고 하고 의미가 어색하지 않은 더 높은 확률의 문장이 선택하는 등 문법적 요소와 의미적 요서를 고려하게 된다. 이렇듯 잘 구축된 언어모형을 활용하면 다양한 자연어의 이해도를 높이는 것이 가능하게 된다고 한다.[15]

생성형 AI의 대표 모델인 Open AI의 GPT의 경우 구글의 트랜스포머 유형을 먼저 살펴볼 필요가 있다. 구글의 트랜스포머 유형은 구글의 번역 서비스를 위하여 개발된 것으로 통계적이고 규칙 기반의 방법을 사용하면서 2016년부터 인공신경망 기술을 기계번역에 적용한(Google Neural Machine Translation, 이하 'GNMT')것에서부터 발달되었다. 문장을 쪼개어 번역하던 방식과 달리 입력 문장의 단어들이 순차적으로 순환신경망에 입력되고 인공 뉴런이 단어의 순서를 반영하여 정보를 기억한 뒤에 다른 언어 문장을 도출한 초기 기계번역의 기본적 구조는 인코더 부분과 디코더 부분으로 구분되어 입력문장의 의미를 일련의 배열로 압축한 뒤, 문장의 맥락이 저장된 내용의 벡터의 값이 디코더의 인공 신경망의 입력 변수가 되어 디코더에서 번역된 단어를 순차적으로 생성하였다.[16] 그러나 이와 같은 방식은 순차적인 특성 때문에 병렬적으

14) <https://www.techtarget.com/whatis/definition/ChatGPT>.
15) 김병필, 각주 9)의 글, 178면 이하.
16) 유원준·안상준, 『PyTorch로 시작하는 딥 러닝 입문』, WikiDocs, 2021. <https://github.com/jerife/DeepLearning-Starting-With-Pytorch#deeplearning-starting-

로 인공 신경망을 학습시키기 어려웠고, 추론의 수행이 이루어지지 못하였으므로 상업적 번역 서비스에서는 한계점을 가져왔다. 이러한 단점을 개선한 것이 구글의 트랜스포머 모형으로, 2017년부터 순환신경망 방식의 단점을 개선하여 어텐션 메커니즘을 통하여 인공신경이 우선순위를 두고 단어를 주목하여 번역하도록 학습시키게[17] 된다. 따라서 인공신경망은 가중치가 부여된 해석을 통하여 병렬처리가 가능해지고 보다 빠른 속도로 인공신경망의 학습이 이루어질 수 있다.[18] 결국 이 트랜스포머 모형의 디코더 부분을 기반으로 한 것이 GPT-2와 GPT-3이라고 할 수 있으며 구글은 이 트랜스포머 기술을 국제적으로 특허 등록을 해 둔 상태이지만 상당수는 오픈 소스로 공개되어 있다고도 한다.[19] 본 연구에서는 생성형 AI에 초점을 두고, '인간이 작성한 것과 같은 텍스트를 만들어 내는 데' 주안을 둔 GPT-3을 중심으로 논의하기로 한다. GPT-4는 엄청난 양의 데이터를 학습한 대형언어모델이고 훨씬 더 방대한 양의 텍스트를 처리할 수 있는 능력을 가지고 있으므로 생성형 AI의 발달을 가속화 할 것임은 명백한 사실이다. 미국 모의 변호사 시험에서도 상위 10%의 성적을 받은[20] 바 있고, 쓰기 및 추론, 코딩 작업에서도 GPT-3.5보다는 높은 성능을 자랑한다. GPT-4의 기술보고서가 모델 개발 방법에 대한 세부 정보를 제공하지 않았으나, 기본적으로 안전성은 GPT-3.5보다 개선되었음을 알 수 있다. 그렇다고 해서 완전히 새로운 패러다임은 아니고 이전 언어 모델의 동일한 제한, 즉 오류와 편견을 추론하고 잘못된 정보를 구성할 수 있는 문제점을

---

with-pytorch>.

17) J. Uszkoreit (2017.) Transformer: A Novel Neural Network Architecture for Language Understanding; 김병필, 각주 9) 논문, 183면 이하 참조. <https://blog.research.google/2017/08/transformer-novel-neural-network.html?m=1>

18) 김병필 변호사는 트랜스포머 모형이 기계번역을 위하여 '인코더-디코더'의 구조를 활용하는 것은 종래의 GNMT모형과 동일하지만 각각의 순환신경망에서 벗어나 다수의 어텐션으로 대체한 것으로 설명한다. 김병필, 각주 9)의 논문, 183면.

19) 구글의 트랜스포머 모형은 2018년의 BERT모형, GPT-2, GPT-3으로 발전하였다. 김병필, 각주 9) 논문 184면.

20) 미국의 변호사 입학 시험인 LSAT에서, GPT-3.5는 149를 받았지만, GPT-4는 163을 받았다. <https://theconversation.com/evolution-not-revolution-why-gpt-4-is-notable-but-not-groundbreaking-201858>.

여전히 가지고 있다. 초기 모델보다는 유해한 콘텐츠를 피하는 데 익숙해 져 있으나 그 진실성과 신뢰성의 문제를 해결하기 위하여 대체의 접근 방식이 요구될 것인지, 아니면 언어모델에서 더 많은 데이터를 투입하면 결국 이런 문제가 해결될 수 있는지에 대한 논쟁은 여전히 존재[21]한다. GPT의 훈련 과정에서 인간 트레이너가 대화를 제공하고 답변에 대하여 순위를 매기기도 하므로 보상의 모델은 최상의 답변을 결정하는 데 도움이 되고 이 과정을 반복하면서 사용자는 향후 대화를 개선하고 추가 피드백을 제공할 수도 있다. 그러나 인간 언어의 복잡성을 완전히 이해하지도 못하고 학습 데이터가 2021년 콘텐츠에 머물러 잘못된 정보를 제공할 가능성이 있다. 이는 GPT가 아직 학습 중이라는 것을 전제하기 때문이다. 입력(input)을 전제로 하여 단어를 생성하기에 답변은 얕아 보이고 학습한 데이터도 2021년도까지의 정보에 그친다. 이는 부적절한 대답을 가능성을 충분히 시사하며 응답 자체가 부자연스러운 것은 차치하고서라도 인용 표기의 출처가 밝혀지지 않는 것도 문제점으로 지적된다. 또한 질문의 순서나 질문 자체에 초점을 맞추기에 다음 질문으로 이동하지 못하고 부적절한 답변이 나올 수 있다. 이러한 기술적인 단점 외에도 윤리적 측면에서도 부정행위, 정보의 사칭, 잘못된 정보의 확산 등에 활용될 가능성도 있다. 또한 코드를 작성할 수 없어 사이버 보안에도 문제점이 지적되고 있다. 전문가들은 맬웨어(malware)를 생성할 수 있는 문제를 제기했고 업데이트 요청을 중지하여 맬웨어 생성에 대한 문제를 해결했다고도 했지만 안전 프로토콜을 우회하는 방법은 얼마든지 찾을 수 있을 것으로 전망된다. 그러나 결국 이런 모든 과정은 훈련 데이터의 편향이라는 문제점보다는 결코 크지 않다. 모델이 가져오는 데이터에 편향이 존재한다면 이는 모델의 출력에 반영될 것이고 불쾌하거나 차별적일 수 있는 언어를 구별하지 못하기 때문이다. 입력 기반 텍스트를

21) M. Scharth (2023). Evolution not revolution: why GPT-4 is notable, but not groundbreaking, The conversation; <https://theconversation.com/evolution-not-revolution-why-gpt-4-is-notable-but-not-groundbreaking-201858>
뿐만 아니라 심지어 실제 시나리오에서 이전 버전에 비해 점진적인 개선일 뿐이며 심사위원들은 GPT-3.5와 GPT-4의 출력을 두고 후자를 선호하는 경우가 61%에 불과하기도 하였다.

사용하면서 잠재적으로 민감한 정보들은 노출될 수밖에 없고 정보를 수집하고 이메일이나 전화번호와 같은 개인정보를 추적하거나 프로파일링(profile) 할 수도 있다. 이러한 경우 정보가 무기한 저장될 수 있는 문제점도 제기되고 있다.[22] 또한 법적으로 불확실성에 따른 문제로서 데이터 수집의 문제와 기술적인 편향의 문제[23], 그리고 개인정보 처리와 개별법에 데이터의 처리가 저촉될 수 있는 가능성[24] 등이 제기 될 수 있다.

2. 발생 가능한 법적 쟁점

특히 민사적으로는, 어떤 형태로 GPT가 콘텐츠를 발생시키더라도, 그 시작은 질문에서 시작되므로 이것은 초기 입력값에 의하여 결정되게 된다. 다시 말해, 새로운 콘텐츠는 어떠한 특별법에도 구속되지 않는 상태로 활용(최초 생성)되게 된다. 그러나 특정 데이터에 접근하게 되면, 데이터는 수정되거나 확장되고 이 과정에서 법적책임이 발생할 여지가 생기게 된다. 사용자가 어떤 종류의 데이터를 입력하던 공급자인 OPEN AI는 이와 관련된 데이터를 제공하게 된다. GPT에 사용된 모든 데이터는 미국에 위치한 제공자가 사용 및 처리하기 위해 제공되고, 처리 범위에 대해서 구체적인 범위가 명시되어 있지 않고 서비스 조건에 따라서 GPT의 제공자는 입력된 데이터의 비공개 사용자를 대상으로 특정한 약속(undertaking)을 하지 않는다. 콘텐츠가 발생함과 동시에 공급자에게 획득되게 되는 것이다. 만약에 사용자의 사업 전략에 관한 데이터

---

22) https://www.techtarget.com/whatis/definition/ChatGPT에서는 악성코드를 맬웨어(malware)로 표기하였음을 밝힌다. 일반적으로 악성코드라고도 번역된다. 맬웨어의 정의에 대하여 다음의 내용을 참조하였다. https://www.techtarget.com/searchsecurity/definition/malware.

23) 김병필, 각주 9)의 논문, 177면 참조.; chatGPT의 훈련 데이터의 편향의 문제는 가장 큰 윤리적인 문제 중 하나로 주장된다. 모델이 가져오는 데이터에 편향이 있다면 이는 모델의 출력에 반영되기 때문이다. 모욕적이라거나 차별적인 언어를 이해하지 못하기 때문에 지속적인 편견을 피하기 위하여 데이터를 꾸준히 검토해야 하나, 다양하고 대표적인 자료를 포함하여 편견을 제어하기 위한 노력이 제시되기도 한다.
https://www.techtarget.com/whatis/definition/ChatGPT

24) T,K, Podrecki (2023). Chat GPT in practice-major legal issues-Personal Data. <https://www.lexology.com/library/detail.aspx?g=33bf4b4f-ffd9-4bf1-bfc1-7c790d86a22f>

가 입력되었다고 가정하자. 사업자가 고객 또는 상대방에 관한 비밀스러운 정보를 입력한다면, 사업자의 상대방에게 무제한으로 열람이 될 수 있는 것이다. 만약 이런 정보들이 의료계약에 의하여 보호되는 데이터의 경우, 계약상의 의무를 위반할 가능성을 가지게 된다. GPT의 활용에 대한 적절한 감독이 필요하다는 것은 주지의 사실이고, 이런 도구를 사용할 수 있는지 여부나 범위까지도 계약자나 실제 사용하는 직원이 어디까지 사용할 수 있을 것인지에 대한 법률이나 지침이 필요하게 된다.[25] OPEN AI사의 제품 및 서비스 개인정보 보호정책에 따르면 사용자가 사용할 때 제공되는 GPT를 포함한 도구(tools)를 포함하여 제공자는 사용자의 정보를 획득하게 된다. 사용자의 정보는 단순한 개인데이터에 국한 되는 것이 아니라 통신, 로그인 세부 정보, 사용에 관한 정보, 분석 및 쿠키 등이 포함된다. 제공된 제품의 서비스의 제공, 유지, 업그레이드, 분석, 새로운 제품 및 서비스 개발 등을 위하여 활용될 수 있다고 공지되고 있고 다양한 종류의 제3자에게도 공개될 수 있고 이 때 개인정보도 포함될 여지가 있다. 데이터가 무제한으로 사용되는 점 이외에도 왜곡되거나 준거법 위반 등의 가능성이 없으리라는 보장이 없으므로 개인정보에 대한 문제는 항상 존재할 수밖에 없게 된다.[26] 결국 민법상의 인격권의 침해에 대한 가능성도 존재하게 되는 것이다. 인격권에 대한 관심이 과거보다 높아진 시점에서[27] 일반적인 인격권의 보호 영역 중 핵심적인 부분은 "사적인 영역의 보호"라고 했을 때

---

25) Traple Konarski Podrecki & Partners (2023). ChatGPT in Practice-major legal issues, <https://www.lexology.com/library/detail.aspx?g=33bf4b4f-ffd9-4bf1-bfc1-7c790d86a22f> 마지막 방문일, 2023. 10. 1. 또한 GPT의 경우 대용량의 데이터를 수집하기 위하여 대부분은 온라인상에서 공개된 텍스트를 웹스크래핑하여 수집하고 구글 북스에 올라온 도서나 ArXiv에 올려진 논문이나, 컴퓨터 프로그램의 소스코드, 게시판에 올려진 이용자들의 질문과 답변, 유튜브의 자막 등의 자료가 모두 망라된다. 인간 작업자에 의하여 일일이 수집될 수도 있지만 bot으로 불리는 봇 또는 크롤러(crwaler)와 같은 스크래핑 도구를 활용하여 자동적으로 다운되어 활용된다. 위와 같은 인공지능 학습 데이터의 수집은 이미 대다수의 인공지능 학습을 위하여 수집되어 있는 것들도 많다. 정보통신망법 제48조 제1항에 금지된 행위가 아닌 지 확인해 보면서, 서비스 제공자가 설정한 스크래핑의 금지범위를 위반하여 수집한 경우에는 위법한 행위로서 처벌의 대상이 될 수 있다.

26) 위의 Traple Konarski Podrecki & Partners-Personal Data 부분 참조.

27) 윤나리, "성명권의 내용 및 인격권에 기한 금지청구권의 인정요건에 관한 고찰", 법조(제72권 4호), 2023., 507면.

의료정보, 일기장과 같은 개인적인 기록이나 정보마저도 기밀성의 보호라는 범주에서 자기보호의 범주에서 인정되는 영역으로 포섭될 수 있을 것이다. "자기보호"라는 규정은 명문으로 없으나 미국 연방대법원의 판례에 의하여 형성이 된 "프라이버시권(Privacy right)"에서 도출되는 사생활의 보호를 의미한다.[28)]

또한 미묘한 상황에 대한 인식도 가능하게 된 GPT-4의 등장으로 인간과 유사한 텍스트를 이해하고 생성할 수 있는 단계에서의 상황인식의 정도가 높아지고 기존의 편견에 대한 문제점도 개선이 될 것으로 기대되고 있다. 이는 복잡한 쿼리[29)]에 대한 깊은 이해와 다양해진 다국어 지원, 심지어 사용자가 표현하지 않은 암시적인 쿼리까지도 유추하는 기능을 제공할 것이라는 기대까지 불러오고 있다. 의료분야에서도 이러한 기술이 환자와 의사와의 의사소통을 촉진할 수 있는 잠재력을 지니고 있다는 평가를 받고 있다. 실제로 위장병학(gastroemterology)의 경우, 의사와 환자의 소통의 도구로서 의료상담의 미래를 엿볼 수 있다고까지 평가되기도 한다. 성공적인 의료서비스의 제공이라는 점에서 환자의 요구나, 두려움, 기대를 이해하는 방향으로 촉진 될 경우 의사가 환자의 상황에 맞추어 고유한 접근방식을 조정할 수도 있게 된다. 소화기 질환은 복잡한 특성으로 인하여 시간 제약이 있는 상황에서는 모든 의사가 일관되게 제공할 수 없는 자세한 설명이나 지침이 필요한 경우가 많아서 GPT-4가 이러한 환자와 의사간의 의사소통을 향상시키고 더 나은 의료결과를 촉진할 수 있으리라 기대되는 것이다. 그 과정은 의사가 질병의 성격, 치료의 선택지, 잠재적인 부작용 및 예후를 명확하게 전달하고 환자는 자신의 건강에 대한 결정에 적극적으로 참여하는 방법으로서 이루어진다.[30)] GPT-4를 통하여 이

---

28) 이부하, "일반적 인격권에 대한 헌법적 고찰-독일 연방헌법재판소 판례 고찰을 통한 일반적 인격권의 구체화-", 미디어와 인격권(제8권 3호), 2022., 141면 이하.

29) 쿼리에 대한 개념은 일반적으로 데이터베이스 테이블이나 테이블의 조합에서 데이터나 정보를 요청하는 것을 의미한다. <https://support.microsoft.com/en-us/office/introduction-to-queries-a9739a09-d3ff-4f36-8ac3-5760249fb65c>.

30) D.A. Drossman, J Ruddy (2020). Improving Patient-Provider Relationships to Improve

러한 데이터를 기반으로 환자의 구체적인 상태에 따라서 생활방식의 변화, 식이요법, 약물 사용에 대한 맞춤형 조언을 제공할 수 있게 된다. 맞춤형 권장 사항은 환자의 순응도를 높이고 뿐만 아니라 GPT-4의 도입으로 일반적인 환자 문의에 대한 응답도 효율성을 높일 것으로 기대되어 진다.[31] 결과적으로 이러한 시스템은 위장병 전문의에게 상당한 시간을 확보해 주겠지만[32] 문제는 이러한 GPT와 같은 인공지능 모델을 활용하면서 매우 민감한 개인 데이터의 침해의 우려를 발생할 수 있다. 환자와 인공지능의 상호작용을 처리한다는 것은 GPT-4를 사용하는 모든 시스템이 무단으로 정보에 접근하는 것과 오용하는 것에 대해 엄격한 보안 프로토콜을 설계해야 함을 의미한다.[33] 그렇다면 현실적으로 환자의 개인정보가 매 단계에서 안전하게 처리될 수 있는 프로세스가 갖추어져야만 한다는 것이다.

III. 생성형 AI의 활용과 개인정보보호

1. 개인정보의 개념과 보호범위

우선 개인정보로서 보호받기 위해서는 먼저 개인정보 보호법 제2조 제1항에서 규정하는 개인정보에 해당 되어야 한다. 법 제2조에서는 "개인정보"란 '살아 있는 개인에 관한 정보로서 다음 각 목의 어느 하나에 해당하는 정보'를 말한다. "가. 성명, 주민등록번호 및 영상 등을 통하여 개인을 알아볼 수 있는 정보 나. 해당 정보만으로는 특정 개인을 알아볼 수 없더라도 다른 정보와 쉽게

Health Care, Clinical Gastroenterology and Hepatology. Vol. 18 Issue 7; 1417-1426.;S. Vahdat, L. Hamzehgardeshi, S. Hessam, Z. Hamzehgardeshi (2014). Patient involvement in health care decision making: a review. Iran Red Crescent Med Journal. Vol. 16 Issue; e12454.

31) S. D. Brown.(2020). The Ethics of Medical AI and the Physician-Patient Relationship. Cambridge Quarterly of Healthcare Ethics, Vol.29 Issue 1.; 118-121.

32) L. Eliot. 각주 7).

33) J. Shaw. (2022). Emerging Paradigms for Ethical Review of Research Using Artificial Intelligence. American Journal of Bioethics Vol.22 Issue5; 42-44.

결합하여 알아볼 수 있는 정보. 이 경우 쉽게 결합할 수 있는지 여부는 다른 정보의 입수 가능성 등 개인을 알아보는 데 소요되는 시간, 비용, 기술 등을 합리적으로 고려하여야 한다. 다. 가목 또는 나목을 제1호의2에 따라 가명처리 함으로써 원래의 상태로 복원하기 위한 추가 정보의 사용·결합 없이는 특정 개인을 알아볼 수 없는 정보(이하 "가명정보"라 한다) 가명처리란 개인정보의 일부를 삭제하거나 일부 또는 전부를 대체하는 등의 방법으로 추가 정보가 없이는 특정 개인을 알아볼 수 없도록 처리하는 것을 말한다."고 규정되어 있다. 제2조 제1호 본문의 규정된 공통요건을 충족하면서 동호 가목 내지는 다 목에서 규정하는 어느 하나의 개별요건을 충족하면 되는데, 의료정보는 우선 "정보"에 해당함을 판단하는 바, 정보라는 것은 범위가 통일되어 있지는 않으나, 지능정보화기본법 제2조 제1호에서 "정보를 광(光) 또는 전자적 방식으로 처리되는 부호, 문자, 음성, 음향 및 영상 등으로 표현된 모든 종류의 자료 또는 지식"으로 정의하면서 처리방식의 제한을 두고 있고, 개인정보보호위원회에서는 "개인정보에 있어 정보의 형태나 내용, 처리방식 등에는 특별한 제한이 없고, 주관적 평가나 부정확한 정보, 허위의 정보까지도 경우에 따라 개인정보에 해당할 수 있다"고 설명하는 바, 이 지침은 전자적 파일은 물론이고 인쇄물, 서면 등 모든 형태의 개인정보파일을 운용하는 개인정보처리자에게 적용됨을 명시하고 있어서 모든 형태의 정보가 개인정보에 해당함을 전제로 한다. 개인정보보호법은 데이터와 정보를 개별적으로 구분하지 않는 것으로 보이며 정보의 개념에 포섭된다. 또한 개인 정보는 "개인(個人)"에 의한 것으로 민법상의 사람에 해당하는 자연인을 의미하고, 살아있는 개인에 대한 정보를 의미한다고 하여 사망자의 정보라 할지라도 생존한 유족과의 관계를 알 수 있다면 유족의 개인정보에도 해당된다고 한다.[34] 이렇게 보았을 때 의료기관에서 수집하고 관리하는 모든 정보는 개인정보법상의 정보의 범위에 포함된다고 할 것이다. 특히 의료법제 19조 및 형법 제317조 제1항에서 규정하는 의료법상의 '비밀유

34) 고학수 외, 김은수,『인공지능 시대의 개인정보 보호법』박영사, 2022., 7-9면.

지의무'의 대상 또한 살아있는 개인의 비밀만을 대상으로 하지 않으므로[35] 개인정보보호법의 보호를 받는다 할 것이다.

빅데이터, 인공지능 등의 기술과 관련한 개인정보의 활용은 개인정보의 보호와 활용이라는 두 가지 목적을 달성하고자 '비식별조치'에 대한 논의가 먼저 시작되었다. '비식별조치'란 연구목적으로 데이터를 공개하더라도 공개의 대상이 되는 데이터에 포함되어 있는 개인의 식별정보를 삭제하거나 임의의 숫자로 대체하여 개인을 식별하지 못하도록 조치를 하였는데 이것만으로는 개인의 프라이버시를 완전히 보호하기는 어려워 보다 가명처리된 정보라 할지라도 공중에 제한 없이 공개되어서는 안 되는 것을 이른다. 2020년 '개인정보 보호법'과 '신용정보법'을 개정하면서 '가명정보 처리'에 관한 특례를 도입하게 되었다.[36] 따라서 현재는 가명처리를 비롯한 비식별 조치에 대한 사항을 포함하는 개정이 이루어진 상태이다. 이는 2020년 8월부터 시행되고 있다.[37]

2. 개인정보 보호와 보건의료데이터 관련 법제

보건의료와 관련 된 개인정보는 의료법 제19조[38], 약사법 제87조[39], 건강검진기본법 제23조[40], 보건의료기본법 제13조[41], 암 관리법 제49조[42], 장

35) 위의 책, 9면.

36) 비식별 조치에 대한 연구는 고학수 외, 각주33)의 책 51면-54면 내용 참조하여 정리하였다.

37) 고학수 외, 앞의 책, 5면.

38) 제19조(정보 누설 금지) ① 의료인이나 의료기관 종사자는 이 법이나 다른 법령에 특별히 규정된 경우 외에는 의료·조산 또는 간호업무나 제17조에 따른 진단서·검안서·증명서 작성·교부 업무, 제18조에 따른 처방전 작성·교부 업무, 제21조에 따른 진료기록 열람·사본 교부 업무, 제22조제2항에 따른 진료기록부등 보존 업무 및 제23조에 따른 전자의무기록 작성·보관·관리 업무를 하면서 알게 된 다른 사람의 정보를 누설하거나 발표하지 못한다. <개정 2016. 5. 29.> ② 제58조제2항에 따라 의료기관 인증에 관한 업무에 종사하는 자 또는 종사하였던 자는 그 업무를 하면서 알게 된 정보를 다른 사람에게 누설하거나 부당한 목적으로 사용하여서는 아니 된다. <신설 2016. 5. 29.>

39) 제87조(비밀 누설 금지) ① 약사·한약사는 이 법 또는 다른 법령에 규정된 경우 외에는 의약품을 조제·판매하면서 알게 된 타인의 비밀을 누설하여서는 아니 된다. <개정 2007. 10. 17.> ② 제47조의3제2항에 따라 의약품 품목허가를 받은 자·수입자 및 의약품 도매상 등의 영업에 관한 비밀을 업무상 알게 된 자는 그 비밀을 타인에게 누설하거나 업무목적 외의 용도로 사용하여서는 아니 된다. <신설 2007. 10. 17., 2016. 12. 2.>

40) 제23조(비밀누설의 금지) 이 법에 따른 국가건강검진 업무에 종사하는 자 또는 종사하였던

기 등 이식에 관한 법률 제31조[43], 생명윤리 및 안전에 관한 법률 제18조[44] 및 제19조[45], 인체조직안전 및 관리 등에 관한 법률 제22조[46] 등에서 다루어

자는 그 업무상 알게 된 비밀을 정당한 이유 없이 누설하여서는 아니 된다.

41) 제13조(비밀 보장) 모든 국민은 보건의료와 관련하여 자신의 신체상·건강상의 비밀과 사생활의 비밀을 침해받지 아니한다. [전문개정 2010. 3. 17.]

42) 제49조(개인정보의 목적 외 사용 금지) 이 법에 따라 암 관리사업에 종사하거나 종사하였던 사람은 「개인정보 보호법」 제18조제2항에 따른 경우를 제외하고는 업무상 알게 된 개인정보를 타인에게 제공 또는 누설하거나 목적 외의 용도로 사용하여서는 아니 된다. <개정 2011.3.29>

43) 제31조(비밀의 유지) ① 국립장기이식관리기관·등록기관·뇌사판정기관(제16조제5항에 따른 의료기관을 포함한다)·이식의료기관(제25조제3항 단서에 따른 의료기관을 포함한다)·뇌사판정대상자관리전문기관·장기구득기관에 종사하는 사람으로서 대통령령으로 정하는 사람은 이 법에서 특별히 규정한 경우 외에는 해당 장기등기증자 등의 등록, 뇌사판정대상자 관리, 장기구득 또는 장기 등의 적출·이식과 관련된 업무를 담당하는 사람 외의 사람에게 다음 각 호의 어느 하나에 해당하는 행위를 하여서는 아니 된다. 1. 장기등기증자와 적출한 장기 등에 관한 사항을 알려주는 행위 2. 이식대상자와 이식한 장기 등에 관한 사항을 알려주는 행위 3. 장기 등 기증희망자 및 장기 등 이식대기자에 관한 사항을 알려주는 행위 ② 다음 각 호의 어느 하나에 해당하는 경우에는 제1항을 적용하지 아니한다. <개정 2021.12.21.> 1. 범죄수사를 위하여 수사기관이 장기 등의 적출 또는 이식에 관련된 자료를 요청한 경우 2. 재판과 관련하여 법관이 장기 등의 적출 또는 이식과 관련된 자료의 제출명령을 한 경우 3. 제6조제2항제1호에 따라 장기 등 기증희망자임을 표시하기 위하여 해당 증명서의 발급기관에 장기 등 기증희망자의 명단을 제공하는 경우 4. 장기기증 홍보사업 등 공익 목적을 위하여 장기등기증자, 장기등기증자의 유족 또는 장기 등 이식대기자 본인이 정보제공에 동의한 경우.

44) 제18조(개인정보의 제공) ① 인간대상연구자는 제16조제1항에 따라 연구대상자로부터 개인정보를 제공하는 것에 대하여 서면동의를 받은 경우에는 기관위원회의 심의를 거쳐 개인정보를 제3자에게 제공할 수 있다. ② 인간대상연구자가 제1항에 따라 개인정보를 제3자에게 제공하는 경우에는 익명화하여야 한다. 다만, 연구대상자가 개인식별정보를 포함하는 것에 동의한 경우에는 그러하지 아니하다.

45) 제19조(기록의 유지와 정보의 공개) ① 인간대상연구자는 인간대상연구와 관련한 사항을 기록·보관하여야 한다. ② 연구대상자는 자신에 관한 정보의 공개를 청구할 수 있으며, 그 청구를 받은 인간대상연구자는 특별한 사유가 없으면 정보를 공개하여야 한다. ③ 제1항 및 제2항에 따른 기록·보관 및 정보 공개에 관한 구체적인 사항은 국가위원회의 심의를 거쳐 보건복지부령으로 정한다.

46) 제22조(비밀의 유지) ①등록기관·조직기증지원기관·조직은행·조직이식의료기관 또는 조직의 기증·관리 및 이식 관련 업무에 종사하는 자로서 대통령령이 정하는 자는 이 법에 특별히 규정한 경우를 제외하고는 당해 조직기증자 등의 조직 채취·이식과 관련된 업무를 담당하는 자 외의 자에게 다음 각 호의 1에 해당하는 사항을 알려주어서는 아니 된다. <개정 2014. 1. 28., 2014. 3. 18., 2016. 2. 4.>

1. 조직기증자와 채취한 조직에 관한 사항, 조직기증자의 병력 및 투약이력의 조사결과에 관한 사항, 조직기증자 및 잠재적 조직기증자의 의무기록에 관한 사항, 장기구득기관에서 연계 받은 뇌사추정자 및 뇌사판정대상자에 관한 사항 2. 이식대상자와 이식한 조직에 관한 사항 3. 조직 기증희망자에 관한 사항 ②다음 각 호의 1에 해당하는 경우에는 제1항의 규정을 적용하지 아니한다. <개정 2014. 1. 28.> 1. 범죄수사를 위한 수사기관이 조직

지고 있다. 개인정보 보호법은 개인정보의 보호와 규제에 관한 일반법으로서 법 제6조에서 '다른 법률에 특별히 규정이 있는 경우를 제외하고는 개인정보 보호법에서 정하는 바에 따르도록' 규정되어 있다. 보건 의료 관련하여 전자처방전이나 전자의무기록에 저장된 개인정보가 침해된 경우에 처벌 규정을 두고 있는데 종이로 된 처방전에 대해서는 명시적으로 개인정보 침해에 대하여 규정이 없다. 이럴 경우, 종이처방전에 대하여 보충적으로 개인정보보호법을 적용하여 처벌이 가능할 것이다.[47]

생성형 AI의 활용이 보편화 될 경우, 앞에서 살펴보았듯이, 많은 정보가 무작위로 수집되고 활용되게 된다. 사실 어디서부터 어떻게 수집될지 확정할 수도 없는 경우도 많을 것이다. 해외 문헌에서는 FDA는 전통적으로 웹상의 의료정보를 규제하려고 하지 않았고, 많은 임상의들이 하루에도 자주 구글(GOOGLE)사의 정보를 참고한다는 것을 들면서 인류에게 알려진 거의 의료상태는 매우 다양할 것으로 예측되는데 결국 GPT-4와 같은 프로그램의 사용으로 인한 당황스러운 사건 혹은 연구 또는 자체의 규제들로 인하여 FDA가 개입하게 될 필요성이 증대될 것이라고도 전망하고 있다.[48]

구체적으로 개인정보와 관련해서 상정할 수 있는 문제로 개인정보 재식별과 관련된 사례를 들 수 있다. 인공지능의 생래적인 특성 중의 예측불가능성과 지속적인 학습능력에서 파생되는 불확실성은 인공지능의 고유특성으로 해석된다.[49] 의료과오소송에서 원고와 피고를 정하려면 그에 맞는 책임의 주체가 분명해야만 한다. 이것이 민사법의 법리인데, 현재의 생성형 AI의 경우, 자료의 근거를 제시하지 않는다. 또한 빅 데이터를 활용하여 머신러닝을 거치므로

등의 채취 또는 이식과 관련된 자료를 요청한 경우 2. 재판과 관련되어 법관이 조직 등의 채취 또는 이식과 관련된 자료를 요청한 경우 3.「장기등 이식에 관한 법률」에 따라 장기 등 기증을 목적으로 요청한 경우.

47) 이정수,『IT시대 개인정보』, 법률신문사, 2023. 18면.

48) Isaac Kohane, Carey Goldberg, Peter Lee, "Safety Fist",『The AI revolution in Medicine: GPT-4 and Beyond』, Pearson Education, Inc, 2023, 256-257p.

49) David F. Llorca, Vicky Charisi, Ronan Hamon, Ignacio Sánchez, Emilia Gómez, "Liability Regimes in the Age of AI: a Use-Case Driven Analysis of the Burden of Proof",『Journal of Artificial Intelligence Research』, Vol 76, 2023, 614p.

개발자가 기본적으로 설정하였던 처리의 범위를 넘어설 수도 있고 결과값은 예측이 되지 않으니 누구의 과실이 손해를 가져왔다고 보기도 힘들다. 다만, 개인의 정보가 누출된 경우 애초에 비식별 정보였으나 재식별로 인하여 특정이 가능하게 되어 문제[50]가 될 경우, 이를 처리한 병원이나 의원, 개인정보처리자에게 책임을 묻게 될 가능성을 배제할 수 없다. 매사추세츠 주정부는 단체보험위원회를 통하여 개인정보를 식별할 수 없을 정도로 삭제한 뒤에 연구목적으로 의료정보를 공개하였는데, 당시 대학원생이던 Sweeney는 유권자 명부와 의료정보를 대조한 결과 매사추세츠 주지사인 William Weld의 의료기록을 재식별 해 낸 것이다. 간단한 대조만으로 단체보험위원회의 의료정보에서 특정인을 재식별 해 낼 수 있었다는 것은 놀라움을 안겨준다. 이것은 적절한 비식별 조치가 이루어지지 않을 경우의 위험성에 대한 경각심을 일깨운 중요한 선례라고 생각된다.[51] 이 사건으로 인하여 해당 데이터의 집합과 결합될 수 있는 보조정보도 고려하여 의료정보가 공개되어야 함을 시사한다.[52]

3. 개인정보 관련 법제

가. 유럽의 GDPR(General Data Protection Regulation, 'GDPR')

비교법적으로 개인정보의 발달에서 살펴보아야 할 개념으로서 유럽연합의 개인정보에 관한 일반규정인 GDPR[53]을 들 수 있다. 2018년 시행되어 오늘에

---

50) 고학수, 『개인정보 비식별화 방법론』, 박영사, 2017., 70면 이하; 고학수 외, 각주 33)의 책, 48면 이하.

51) 김병필, 「개인정보 위험기반 비식별 조치와 가명처리」, 서울대학교 대학원 법학석사학위 논문, 2021, 10면.

52) 고학수 외, 각주 33)의 책, 49면.

53) Art. 1 GDPR Subject-matter and objectives ; This Regulation lays down rules relating to the protection of natural persons with regard to the processing of personal data and rules relating to the free movement of personal data. This Regulation protects fundamental rights and freedoms of natural persons and in particular their right to the protection of personal data. The free movement of personal data within the Union shall be neither restricted nor prohibited for reasons connected with the protection of natural persons with regard to the processing of personal data. 본 규정은 개인정보의 처리와 관련된 자연인의 보호와 관련된 규칙과 개인정보의 자유로운 이동과 관련된 규칙을 규정하고 있다. 본 규정은 자연인의

이르고 있는데 이전의 1995년 개인정보보호지침(Directive 95/46/EC)과는 달리 GDPR은 각 회원국의 국내 입법이 없다 하더라도 직접적인 법적 구속력을 가지게 된다.[54] GDPR의 역외적용조항에 의하여 우리나라의 개인정보처리 주체들도 적용을 받을 가능성이 있고 우리나라의 개인정보 보호법 또한 유럽연합의 GDPR에 상응하는 개인정보 보호수준으로 적정성을 인정받을 수 있는 방향으로 개정되었다.[55] 우리나라의 개인정보 보호법의 규정체계에 대하여 개인정보를 적법하게 처리하기 위해서는 정보주체의 동의를 얻고 기타 법률에 규정된 적법요건을 충족해야 하는 것으로 틀을 둔 것을 두고 유럽연합의 GDPR과 매우 유사하다고 평가하는 견해도 있다.[56] GDPR의 제6조 제2항 (b)에서는 개인정보 처리의 적법성의 여부는 GDPR의 목적과 전체내용을 아우르는 맥락에서 고려되도록 하면서 해당 개인정보의 처리가 계약이 추구하는 목적에 부합하는지와 동일한 목표를 달성하고자 하는 다른 수단에 비하여 덜 침익적인 수단인지에 기반하여 평가하도록 하고 있다.[57] 특히 GDPR의 제4조 제1항에서는 식별된 또는 식별가능한 자연인과 관련된 "일체의 정보"를 개인정보로 정의하는 만큼 개인정보를 널리 인정하고 보호하려는 입법취지를 알 수 있다. 개인정보의 해석에 있어서 정보의 성질, 내용, 형식에서 설명하고 있으나 본 연구에서 논의되는 의료정보는 그 성질이나 내용이 의료관계에서 생성되고 보관되는 일체의 의료정보로서 파악이 되는 정보군으로 상정이 가능하

기본권과 자유, 특히 개인정보보호에 대한 권리를 보호한다. 개인정보의 처리와 관련하여 자연인의 보호와 관련된 이유로 연합 내에서 개인정보의 자유로운 이동을 제한하거나 금지하지 않는다.

54) Art. 99 GDPR Entry into force and application ; This Regulation shall enter into force on the twentieth day following that of its publication in the Official Journal of the European Union. It shall apply from 25 May 2018. <https://gdpr-info.eu/art-99-gdpr/>

55) 고학수 외, 각주33)의 책, 28-29면.

56) 이해원, "개정 개인정보 보호법의 민사법적 쟁점", 법조(제72권 3호), 2023., 90면 이하.

57) 이해원, 앞의 글, 90-91면. 이 글에서 GDPR의 해석에 대하여 유럽 개인정보보호 이사회(European Union Data Protection Board, 'EDPB')의 2018년 가이드라인에서의 GDPR의 해석을 소개하고 있다. 이해원 교수는 EDPB는 GDPR 제6조 제1항 (b)는 "최대한 엄격하게 해석 및 적용"되도록 하고, "해당 조항이 정보주체의 동의(GDPR 제6조 제1항(a))를 회피하는 수단으로 사용됨을 경계"한다고 평가하고 있다.

므로 수치 등으로 표현되는 객관적 정보는 당연히 포함이 된다고 보인다. 더 나아가 특정 자연인에 대한 평가 내지 의견을 담은 주관적 정보 모두도 일체의 정보로서 개인정보에 해당될 수 있다고 보아야 할 것이다.[58] 심지어 GDPR은 개인정보가 반드시 사생활에 관한 것일 필요가 없다[59]고까지 밝히고 있으므로 의료정보는 모두 개인정보에 포함하여 보호하는 것이 법 위반의 상황을 피할 수 있는 해석이라 판단된다.

나. 미국의 "건강보험이전과 책임에 관한 법(Health Insurance Portability and Accountability Act, 'HIPPA')

미국은 연방과 주의 이중적 법체계를 가지고 있는데 프라이버시 또는 개인정보보호에 관하여도 연방법과 개별 주법이 산재한 상황이다. 연방법 차원에서 모든 분야의 개인정보를 포괄적으로 규율하는 일반법은 없고, 분야별 개별 입법으로 규제하고 있다. 의료정보 분야에서는 "건강보험이전과 책임에 관한 법"의 프라이버시 규칙이 비식별화의 기준을 설정하고 있다.

일반적인 논의의 선상에서 개인정보는 "개인적으로 식별가능한 정보"라고 전제하고 개인정보란 기관이 보유하고 있는 개인에 대한 정보를 의미한다. 대표적으로 개인의 신분을 구별하거나 추적할 수 있는 용도로 이용될 수 있는 정보, 개인과 연결되거나 연결될 수 있는 기타정보(의료, 교육, 그리고 고용에 대한 정보) 이러한 정의에 의하면 규제 대상은 식별가능성이 있는 정보로 제한되고 이러한 비식별 정보는 개인정보보호에 대한 법의 적용을 받지 않는다. 자연스럽게 미국은 비식별화에 관심이 집중되어 가이드라인이 발표되었다. 모든 비식별화 데이터에는 언제나 재식별의 위험이 내재되어 있다는 점을 기본적인 전제로 하여 개별적인 사례별로 데이터가 수집되고 처리되는 맥락에 따라서 재식별이 될 위험성의 정도가 달라질 것이므로 이러한 차이를 적절히 고려할 필요가 있다고 강조한다.[60]

---

58) 고학수 외, 각주33)의 책, 41면 참조.
59) 고학수 외, 각주33)의 책, 42면.

다. 우리나라의 보건의료 분야의 개인정보 보호

우리 정부는 데이터에 활용을 대비하여 2020년 2월 '개인정보보호법'을 개정하며 가명정보의 개념과 활용에 대한 법적 근거를 신설하였다.[61] 개인정보보호위원회에서는 '가명정보 처리 가이드라인'을 개정하였다. 이는 통계를 작성하거나 연구목적으로 개인정보를 활용할 때, 공익적 기록보존 등을 위해서 개인정보를 가명처리하여 활용할 수 있는 기반을 마련한 것이다. '개인정보보호법', '정보통신망 이용촉진 및 정보보호 등에 관한 법률', '신용정보의 이용 및 보호에 관한 법률'에서 4차 산업혁명 시대의 데이터 활용에 대한 시대적 요구를 반영한 것으로 볼 수 있다. 데이터의 활용을 위한 가명정보의 처리에 대한 이해를 돕기 위한 것으로 보인다. 개인정보처리자가 법에 따른 규정을 준수할 경우에는 가이드라인 미준수를 이유로 처벌받지 않기 때문에 개정 '개인정보 보호법'의 시행에 따른 설명과 구체적 사례를 제공하고 있다. 개인정보보호법 제2조, 제28조의 2의 가명정보 처리 등[62]에서는 제1항 개인정보처리자는 '통계작성, 과학적 연구, 공익적 기록보존 등을 위하여 정보주체의 동의 없이 가명정보를 처리'할 수 있다. 제2항 개인정보처리자는 제1항에 따라 '가명정보를 제3자에게 제공하는 경우에는 특정 개인을 알아보기 위하여 사용될 수 있는 정보를 포함'해서는 아니 된다. 제28조의 3의 가명정보의 결합 제한[63], 제28조의 4 가명정보에 대한 안전조치의무 등[64], 제28조의 5 가명정보 처리 시 금지의무

60) 고학수, 각주 49)의 책, 70면 이하..

61) 이와 함께 2022년 5월 '디지털플랫폼정부 추진방향'을 발표, 가명정보 활용을 기반으로 한 공공데이터 개방과 데이터 기반 행정 정책을 제시함.

62) 제28조의 2(가명정보의 처리 등) ① 개인정보처리자는 통계작성, 과학적 연구, 공익적 기록보존 등을 위하여 정보주체의 동의 없이 가명정보를 처리할 수 있다. ② 개인정보처리자는 제1항에 따라 가명정보를 제3자에게 제공하는 경우에는 특정 개인을 알아보기 위하여 사용될 수 있는 정보를 포함해서는 아니 된다.

63) 제28조의3(가명정보의 결합 제한) ① 제28조의2에도 불구하고 통계작성, 과학적 연구, 공익적 기록보존 등을 위한 서로 다른 개인정보처리자 간의 가명정보의 결합은 보호위원회 또는 관계 중앙행정기관의 장이 지정하는 전문기관이 수행한다. ② 결합을 수행한 기관 외부로 결합된 정보를 반출하려는 개인정보처리자는 가명정보 또는 제58조의2에 해당하는 정보로 처리한 뒤 전문기관의 장의 승인을 받아야 한다. ③ 제1항에 따른 결합 절차와 방법, 전문기관의 지정과 지정 취소 기준·절차, 관리·감독, 제2항에 따른 반출 및 승인 기준·절차 등 필요한 사항은 대통령령으로 정한다.

등[65]이 포함되었다. 제28조의 8에서는 개인정보의 국외 이전[66]까지 규정이

64) 제28조의4(가명정보에 대한 안전조치의무 등) ① 개인정보처리자는 제28조의2 또는 제28조의3에 따라 가명정보를 처리하는 경우에는 원래의 상태로 복원하기 위한 추가 정보를 별도로 분리하여 보관·관리하는 등 해당 정보가 분실·도난·유출·위조·변조 또는 훼손되지 않도록 대통령령으로 정하는 바에 따라 안전성 확보에 필요한 기술적·관리적 및 물리적 조치를 하여야 한다. <개정 2023. 3. 14.> ② 개인정보처리자는 제28조의2 또는 제28조의3에 따라 가명정보를 처리하는 경우 처리목적 등을 고려하여 가명정보의 처리 기간을 별도로 정할 수 있다. <신설 2023. 3. 14.> ③ 개인정보처리자는 제28조의2 또는 제28조의3에 따라 가명정보를 처리하고자 하는 경우에는 가명정보의 처리 목적, 제3자 제공 시 제공받는 자, 가명정보의 처리 기간(제2항에 따라 처리 기간을 별도로 정한 경우에 한한다) 등 가명정보의 처리 내용을 관리하기 위하여 대통령령으로 정하는 사항에 대한 관련 기록을 작성하여 보관하여야 하며, 가명정보를 파기한 경우에는 파기한 날부터 3년 이상 보관하여야 한다. <개정 2023. 3. 14.>

65) 제28조의5(가명정보 처리 시 금지의무 등) ① 제28조의2 또는 제28조의3에 따라 가명정보를 처리하는 자는 특정 개인을 알아보기 위한 목적으로 가명정보를 처리해서는 아니 된다. <개정 2023. 3. 14.> ② 개인정보처리자는 제28조의2 또는 제28조의3에 따라 가명정보를 처리하는 과정에서 특정 개인을 알아볼 수 있는 정보가 생성된 경우에는 즉시 해당 정보의 처리를 중지하고, 지체 없이 회수·파기하여야 한다. <개정 2023. 3. 14.>

66) 제28조의8(개인정보의 국외 이전) ① 개인정보처리자는 개인정보를 국외로 제공(조회되는 경우를 포함한다)·처리위탁·보관(이하 이 절에서 “이전”이라 한다)하여서는 아니 된다. 다만, 다음 각 호의 어느 하나에 해당하는 경우에는 개인정보를 국외로 이전할 수 있다.
1. 정보주체로부터 국외 이전에 관한 별도의 동의를 받은 경우
2. 법률, 대한민국을 당사자로 하는 조약 또는 그 밖의 국제협정에 개인정보의 국외 이전에 관한 특별한 규정이 있는 경우
3. 정보주체와의 계약의 체결 및 이행을 위하여 개인정보의 처리위탁·보관이 필요한 경우로서 다음 각 목의 어느 하나에 해당하는 경우
   가. 제2항 각 호의 사항을 제30조에 따른 개인정보 처리방침에 공개한 경우
   나. 전자우편 등 대통령령으로 정하는 방법에 따라 제2항 각 호의 사항을 정보주체에게 알린 경우
4. 개인정보를 이전받는 자가 제32조의2에 따른 개인정보 보호 인증 등 보호위원회가 정하여 고시하는 인증을 받은 경우로서 다음 각 목의 조치를 모두 한 경우
   가. 개인정보 보호에 필요한 안전조치 및 정보주체 권리보장에 필요한 조치
   나. 인증받은 사항을 개인정보가 이전되는 국가에서 이행하기 위하여 필요한 조치
5. 개인정보가 이전되는 국가 또는 국제기구의 개인정보 보호체계, 정보주체 권리보장 범위, 피해구제 절차 등이 이 법에 따른 개인정보 보호 수준과 실질적으로 동등한 수준을 갖추었다고 보호위원회가 인정하는 경우

② 개인정보처리자는 제1항제1호에 따른 동의를 받을 때에는 미리 다음 각 호의 사항을 정보주체에게 알려야 한다.
1. 이전되는 개인정보 항목
2. 개인정보가 이전되는 국가, 시기 및 방법
3. 개인정보를 이전받는 자의 성명(법인인 경우에는 그 명칭과 연락처를 말한다)
4. 개인정보를 이전받는 자의 개인정보 이용목적 및 보유·이용 기간
5. 개인정보의 이전을 거부하는 방법, 절차 및 거부의 효과

③ 개인정보처리자는 제2항 각 호의 어느 하나에 해당하는 사항을 변경하는 경우에는 정보주체에게 알리고 동의를 받아야 한다.

되어 있다. 이것은 대한민국의 국민의 개인정보가 국외로 이전된 이후에 제3국으로 제공되고 처리위탁 및 보관되는 경우(재이전)이 허용되는 경우를 전제하고 있다. 보호위원회는 일정한 경우에 국외 재이전에 대하여 중지명령을 내릴 수 있는데[67] 이 경우 국외의 재이전에서의 개인정보처리자는 개인정보를 이전받는 자로, 개인정보를 이전받는 자는 제3국에서의 이전받는 자로 보게 된다. 이 조항과 관련하여 국외로 이전된 개인정보의 보호정도가 본 법에서 규정한 수준으로 보호될 수 있을지에 대하여 GDPR과 달리 국외 재이전의 대상에 국제기구가 포함되어 있지 않아서 개선이 필요함이 지적되기도 한다.[68]

그러나 본 가이드라인의 적용 대상은 제3장 제3절의 가명정보 처리에 관한 특례에 근거한 가명정보 처리이기에 개인정보보호위원회가 소관 부처와 공동으로 발간한 개인정보의 가명정보의 처리에 관한 분야별 가이드라인이 있다면 해당 분야의 가이드라인이 우선 적용되게 된다. 따라서 보건의료 데이터의 활용의 경우에는 보건 복지부에서 이미 가이드라인을 제공하여 이것이 먼저 적용되게 된다.[69] 이는 의료기관, 연구자, 기업, 공공기관, 대학교 등 보건의료

---

④ 개인정보처리자는 제1항 각 호 외의 부분 단서에 따라 개인정보를 국외로 이전하는 경우 국외 이전과 관련한 이 법의 다른 규정, 제17조부터 제19조까지의 규정 및 제5장의 규정을 준수하여야 하고, 대통령령으로 정하는 보호조치를 하여야 한다.

⑤ 개인정보처리자는 이 법을 위반하는 사항을 내용으로 하는 개인정보의 국외 이전에 관한 계약을 체결하여서는 아니 된다.

⑥ 제1항부터 제5항까지에서 규정한 사항 외에 개인정보 국외 이전의 기준 및 절차 등에 필요한 사항은 대통령령으로 정한다. [본조신설 2023. 3. 14.]

67) 개인정보 보호법 제28조의9(개인정보의 국외 이전 중지 명령) ① 보호위원회는 개인정보의 국외 이전이 계속되고 있거나 추가적인 국외 이전이 예상되는 경우로서 다음 각 호의 어느 하나에 해당하는 경우에는 개인정보처리자에게 개인정보의 국외 이전을 중지할 것을 명할 수 있다.

1. 제28조의8제1항, 제4항 또는 제5항을 위반한 경우
2. 개인정보를 이전받는 자나 개인정보가 이전되는 국가 또는 국제기구가 이 법에 따른 개인정보 보호 수준에 비하여 개인정보를 적정하게 보호하지 아니하여 정보주체에게 피해가 발생하거나 발생할 우려가 현저한 경우

② 개인정보처리자는 제1항에 따른 국외 이전 중지 명령을 받은 경우에는 명령을 받은 날부터 7일 이내에 보호위원회에 이의를 제기할 수 있다.

③ 제1항에 따른 개인정보 국외 이전 중지 명령의 기준, 제2항에 따른 불복 절차 등에 필요한 사항은 대통령령으로 정한다. [본조신설 2023. 3. 14.]

68) 법 제28조의 11 제2문; 박노형, “개정된 개인정보보호법상 개인정보의 국회이전에 관한 규정의 분석: GDPR을 참조하여”, 고려법학(제109권), 2023., 185면.

데이터를 처리하는 모든 개인정보처리자에게 적용되는 것으로 '보건의료정보'라 함은 「보건의료기본법」 제3조 제6호에 의한 보건의료정보로서 '보건의료와 관련한 지식 또는 부호·숫자·문자·음성·음향·영상 등으로 표현한 모든 종류의 자료로써 광(光) 또는 전자적 방식으로 처리될 수 있는 것'을 의미한다.[70] 보건의료 데이터와 관련하여서는 고유의 식별번호를 활용하여 제3의 신뢰기관이 발급한 랜덤 키값을 결합해 사용하고 비식별화에 초점을 두고 있다. 우리나라는 보건의료 빅 데이터의 통합 플랫폼을 운영하면서 공공 보건의료 데이터의 활용을 통하여 시스템을 갖추고 있다.[71] 우리나라의 보건의료 분야에서의 개인정보 보호법제는 개정 개인정보 보호법의 시행과 함께 데이터 활용에 있어서의 핵심인 가명정보의 활용에 대한 법적근거가 마련되면서 보건의료데이터의 특수성이 고려되었다고 보인다. 보건의료 데이터는 그 분야와 유형, 목적별 세부기준과 절차를 제시하여 현장에서 활용하려 하였을 때 혼란을 방지하고 자료의 오남용을 방지하는 것을 목적으로 하고 있다. 개인정보 보호법의 가명정보 처리와, 결합제한, 안전조치의무, 금지의무 등을 따르면서 가명처리의 원칙에서도 위원회의 수를 기존의 5인 이상 15인 이하에서 5인 이상으로 개정하고 보건의료데이터를 이용한 연구경험이 풍부하고 전문성을 갖춘 자를 구성하도록 하였다. 또한 영상정보에 관한 규정을 개정하면서 영상정보상의 환자번호, 성명, 생년월일 등의 개인정보 일부를 삭제하거나 일부 또는 전부를 삭제하도록 하였다.[72] 매우 꼼꼼하게 개정이 되어 있지만, 가명처리 원

69) 개인정보보호위원회, 「가명정보 처리 가이드라인」, 5면 이하. 2022. 4. <https://data.privacy.go.kr/recsroom/recsroomDtl.html?recsroomSeq=35>

70) 보건의료기본법 제3조(정의) 이 법에서 사용하는 용어의 뜻은 다음과 같다. 1. "보건의료"란 국민의 건강을 보호·증진하기 위하여 국가·지방자치단체·보건의료기관 또는 보건의료인 등이 행하는 모든 활동을 말한다. 2. "보건의료서비스"란 국민의 건강을 보호·증진하기 위하여 보건의료인이 행하는 모든 활동을 말한다. 3. "보건의료인"이란 보건의료 관계 법령에서 정하는 바에 따라 자격·면허 등을 취득하거나 보건의료서비스에 종사하는 것이 허용된 자를 말한다. 4. "보건의료기관"이란 보건의료인이 공중(공중) 또는 특정 다수인을 위하여 보건의료서비스를 행하는 보건기관, 의료기관, 약국, 그 밖에 대통령령으로 정하는 기관을 말한다. 5. "공공보건의료기관"이란 국가·지방자치단체, 그 밖의 공공단체가 설립·운영하는 보건의료기관을 말한다. 6. "보건의료정보"란 보건의료와 관련한 지식 또는 부호·숫자·문자·음성·음향·영상 등으로 표현된 모든 종류의 자료를 말한다.

71) https://hcdl.mohw.go.kr/static/info/introPlatform.

칙에서 “안전한 가명처리 방법이 없을 경우 가명정보로 변환하여 활용이 가능”하도록 하였는데 이는 “안전한 가명처리의 방법이 현재 개발되지 않은 정보의 경우 가명처리 방법이 개발될 때까지는 가명처리 가능 여부를 판단할 수 없으며, 이러한 정보는 정보주체의 동의 하에서만 활용이 가능하다”라고 되어 있다. 이는 안전한 가명처리의 방법이 없다면 “안전한 가명처리방법이 개발될 때까지 유보가 되고 동의에 의해서만 활용이 가능”한 것으로 설명되고 있다.[73] 과연 그렇다면 안전한 가명처리의 방법과 그것이 개발될 때까지 유보가 된다는 것은 현재의 기술수준에서 어느 정도의 가명처리까지 유보가 될 것인가가 일반인으로서는 경계가 분명하지 않다. 정보주체의 경우 어떠한 진료를 받는다고 했을 때 이런 부분에 모두 동의를 할 수밖에 없는 구조인데 그렇다면 안전한 기술적 방법이 제시되지 않는 동안에는 정보주체는 동의를 할 수밖에 없을 것으로 짐작할 수 있다.

또한 보건의료데이터 활용 가이드라인에 의하면 가명정보를 재 제공 할 목적으로 제공받는 것은 금지되고, 가명정보를 최초 제공받을 당시 원 개인정보처리자에게 밝힌 목적이외의 목적으로 처리하게 되는 경우에는 내부 활용절차에 준하여 심의위원회의 심의를 받아야 하는데 이 경우 원 개인정보처리자에게 내용을 고지할 것을 권장하고 있다. 보건의료데이터의 활용에 있어 매우 세심하게 여러 부분에 법적의무를 부과하였으나 과연 이것이 실제 의료계에서 제대로 적용될 수 있을지는 의문을 제기하지 않을 수 없다. 실제 가명정보의 처리 절차는 내부적으로 의사가 데이터 활용신청서를 작성하여 데이터 담당부서에 활용신청을 하면 심의위원회에서 승인할 것인지의 여부를 심의하고 가명이나 익명처리의 방법 및 활용환경을 결정하고 이 결정에 따라서 병원의 데이터 부서에서 실시하게 된다. 이 경우 데이터 활용에 대한 신청을 제출했던 의사는 직접 가명처리를 실시하지 않고, 데이터 담당부서에서 가명처리를 실시하도록 하는 것[74]

---

72) 보건복지부, 개인정보보호위원회, “보건의료데이터 활용 가이드라인”, 2022. 12., 12면 이하.
73) 보건복지부, 개인정보보호위원회, 위의 가이드라인, 7면.
74) 데이터 처리에 관한 과정은 보건복지부, 개인정보보호위원회, 위의 가이드라인, 16면 이하의 도표와 활용 예시를 참조.

은 매우 바람직하나, 이런 과정을 통하여 보건의료 데이터를 활용하려면 연구를 통해 얻은 데이터를 대체로 짧은 주기의 환자 진료에 활용하거나 즉각적인 상담 등의 진료행위에 적용하기에는 실질적인 어려움이 있을 것으로 전망된다.

4. 생성형 AI의 활용과정에서의 개인정보 침해에 대한 제언

우리의 개인정보 보호법과 보건의료데이터의 활용 가이드는 개인정보의 보호와 과학연구 및 통계목적으로 개인정보가 활용될 경우, 비식별조치 등이 적용 될 수 있도록 매우 섬세하게 규정되어 있는 편이다. 비교법적 기준으로 볼 때에도 유럽의 GDPR과도 견주어 개인정보의 보호범위가 넓고[75] 개인정보의 재식별 가능성을 고려하여 비식별화 조치에도 많은 정성을 들였다. 우리 법제가 4차 산업혁명을 앞두고 많은 시간 공들여 개인정보 보호 법제를 위하여 노력해 온 산물이다. 그런데 2023년 생성형 AI의 급성장은 기술의 급속한 배포와 규제 및 감독의 부족으로 인해 다소 위험한 것으로 HIMSS컨퍼런스[76]에서 전문가들에 의해 지적되고 있다. 생성형 AI가 미래의 의료를 개선할 엄청난 잠재력을 지녔다는 지지자들과, 생성형 AI의 윤리적 어려움을 지적하는 견해도 만만치 않다. 마이크로소프트 사의 피터 리(Peter Lee) 부사장은 기조연설을 통해 생성형 AI가 다양한 활용 사례에서 초기 성공을 보이고 있다고 밝히면서 의사들에게는 새로운 의료 사례를 해석하거나 표준화 된 방식으로 환자와의

---

75) GDPR에서 개인정보의 개념을 특정 자연인의 내밀한 사생활과 관련한 정보뿐만 아니라, 정보 그 자체로서 개인의 사회적, 경제적 활동 전반에 걸친 일체의 정보를 개인정보에 해당할 가능성을 제시하고 있다. 고학수 외, 앞의 책, 42-43면. 특히 문자 형태가 아닌 영상이나 화상, 음성 등 모든 형식의 정보를 아우르고 있다. 우리나라의 보건의료 데이터도 지식, 부호, 숫자, 문자, 음성, 음향, 영상 등 표현된 모든 종류의 자료를 포괄하고 있기에 보호의 범위는 유사한 것으로 보인다. 보건복지부, 개인정보보호위원회, 위의 가이드라인, 8면.

76) 시카고에서 열린 2023 HIMSS는 Healthcare Information and Management Systems Society의 약자로서 글로벌 건강 생태계의 전반에 걸친 수만 명의 전문가들이 차세대 개척지인 디지털 헬스케어를 다루기 위하여 모이는 영향력 있는 의료 이벤트이다. <https://www.himss.org/news/coming- chicago-2023-himss-global-health-conference-exhibition#:~:text=Registration%20is%20o pen%20for%20HIMSS23,at%20McCormick%20Place%20Convention%20Center.>.

대화를 재작성 할 수 있다고 설명했다.[77] 구글도 특정 고객 그룹이 직접 탐색할 수 있는 Med-PaLM2라는 자체의 대규모 언어모델을 만들고 있는데, 이 모델은 의료 데이터에 대한 특별한 교육을 받았으므로 방대한 양의 의료 정보를 선별하고 이해할 수 있다. 구글에서는 이 프로그램이 환자를 대면하는 환경에서 사용되는 것은 아니고 데이터를 분석하여 복잡한 질병을 진단하고 기록을 작성하거나 환자 포털의 컨시어지 역할을 하게 될 것이라고 밝혔다. 이들이 추구하는 것은 생성형 AI의 활용을 통하여 의료 기록을 간소화하고 의사로 하여금 메모를 기록하는 시간들을 절약시켜 피로감을 줄여주는 것이라고 주장한다.[78] 앞서 위장병학에서 환자들에게 단순한 설명을 반복하느라 의사가 허비할 수 있는 시간을 생성형 AI가 줄여줄 수 있다는 것과 같은 맥락이다. 이렇게 실제 의료현장에서 실시간으로 환자와 의사가 소통하는 데에 생성형 AI가 활용된다면, 직접 생성형 AI가 인간 환자의 데이터와 접촉할 수밖에 없고 우리가 일반적으로 법제화 해온 개인정보의 가명처리 등의 절차는 생성형 AI의 활용 환경에서는 적용되기 어려울 수 있게 된다.

다시 말해, 이미 진료 등을 통하여 확보된 개인정보의 경우에는 가명처리 등의 비식별조치를 통하여 존재할 경우, 생성형 AI 등이 머신러닝의 과정을 통해서 개인정보 데이터를 흡수한다 하더라도 개인정보의 유출 가능성은 법제로 대비가 가능할 수 있다. 보건의료데이터의 활용 과정이 비록 완전하지 않다고 하더라도[79] 개인정보 혹은 관련 데이터가 활용되는 데에는 신중한 과정이 요구될 수 있다. 그러나 생성형 AI가 직접적으로 의료기의 소프트웨어에 적용되고 클라우드 시스템을 통하여 자료가 공유된다면 일반인 환자가 자신의 입력하는 정보에 대하여 일일이 특정 정보를 지우도록 할 수 있을지 의문이며, 또한

---

77) Rebecca Pifer,'Hurtling into the future': The potential and thorny ethics of generative AI in healthcare, HEALTHCAREDIVE, 기사입력 2023. 4. 21. 마지막 방문일, 2023. 12. 1. <https://www.healthcaredive.com/news/generative-AI-healthcare-gpt-potential/648104/>

78) 위의 기사, 마지막 방문일, 2023. 12. 1.

79) 심의위원회 등에서 보건의료 데이터를 활용할 수 있도록 승인 여부를 검토할 수 있으므로 한 번 더 숙고하여 개인정보가 다뤄질 수 있음에 주목할 때, 개인정보 유출의 가능성이 보다 신중하게 판단될 여지도 있다.

환자와의 상담을 위한 과정에서 활용되는 경우에도 환자 자신이 직접 본인의 의사로 정보를 입력하게 되므로 이것이 유출되지 않으리란 보장이 없게 된다. 따라서 이러한 경우를 방지하기 위해서는 적어도 보건·의료 환경에서 활용되는 생성형 AI의 소프트웨어에 대입되는 질문에 대해서는 그 입력단계에서 정보의 가명처리 등이 이루어지도록 설정하거나 혹은 유사한 정도의 개인정보 보호를 위한 기술적 조치가 필요할 것으로 보인다. 개인정보 보호법이나 보건의료 데이터에 관한 가이드라인 등은 4차 산업혁명의 빅 데이터의 연구적 목적 및 통계적 목적으로의 활용을 위한 목적으로 개정되고 개인정보를 보호하려는 취지에 부합한다. 그러나 빠른 기술의 발달은 이제 앞으로 인류가 마주할 기술의 정도를 예측하는 것을 넘어서고 있음을 주지하게 한다.

IV. 맺는 말

생성형 AI의 출현과 급격한 성장은 비단 우리나라의 의료계뿐만 아니라 전 세계의 의료계에 데이터의 활용과 콘텐츠의 법적 규제에 대한 새로운 질문을 던진다. 훌륭한 기술의 탄생을 목도하면서도 앞으로 어떻게 활용해야 개인정보가 보호되면서 슬기롭게 인류를 위하여 활용될 수 있을지에 대한 고민이 깊다. 데이터가 어떻게 사용되는지, 사용자가 입력한 정보가 어떻게 이해되고 있는지 GPT의 발달단계에 따라 현격한 차이를 보인다. 생성형 AI의 대표격인 GPT는 버전을 거듭하면서 강화학습을 통해 피드백을 받아서 사용자에 의해 스스로 성능이 높아지고 있는 상황이기 때문이다. 자동화된 도구로서 활용되어 데이터의 출처가 불분명하기에[80] 보건의료 환경에서 활용될 경우 한정적인 역할에서 활용되거나 특정한 임무를 반복하여 환자에게 설명을 하는 용도가 아닌 질문에 대한 응답을 하는 경우라면, 잘못된 정보를 환자에게 제공할 우려도 있다. 그러

80) 김윤명, "생성형 인공지능(AI)모델의 법률문제 - Chat GPT 생성물에 대한 논의를 중심으로 -", 정보법학(제27권 1호), 2023., 82면.

나 GPT는 이미 성공적인 런칭을 했고 인류에게 이미 많이 활용되고 있다. 생성형 AI가 의료계에서 활용되게 되면 단순한 설명 등은 환자가 일정한 이해 수준에 도달할 수 있을 때까지 생성형 AI가 담당하도록 설계될 수도 있고, 환자가 직접 내장된 생성형 AI의 소프트웨어에 자신의 증상에 대한 질문을 하여 해답을 얻을 수도 있다. 그러나 다른 한편으로는 의사가 제공한 치료방법에 대한 의문을 품어 법적 분쟁으로 이어질 수도 있고[81], 머신러닝을 거쳐 의료정보를 학습하여 일정한 수준의 지식은 담보된다고 하더라도 생성형 AI 특히 GPT-4의 경우 여전히 자신이 내어놓은 해답에 대한 출처는 밝혀지지 않아 그 내용의 진위를 두고도 문제가 될 수 있다. 우리나라의 개인정보 보호법과 보건의료데이터 활용 가이드는 환자의 개인정보를 보호하기 위하여 가명정보로 처리하는 등으로 비식별조치를 위한 프로세스를 만들어 두었으나, 생성형 AI가 의료기기에 도입되고 적극적으로 활용된다면 가명처리 단계 자체를 얼마나 준수할 수 있을지 현실적으로 미지수이다. 환자 자신이 입력한 정보에 자신의 민감 정보가 포함될 수 있어 이에 관한 기술적인 조치가 선제적으로 이루어지지 않는 이상 데이터 그 자체에 환자 정보가 포함될 수 있고, 개인정보가 해킹 등으로 유출될 경우 빅 데이터와 함께 섞이게 되면 생성형 AI는 비식별조치가 된 정보와 아닌 정보를 구별하지 못할 것이므로 이에 관한 문제는 인격권의 침해로 이어질 가능성이 높다. 현재의 개인정보 보호법제는 개인정보가 개인정보처리자에 의하여 관리될 경우 프라이버시를 지킬 수 있도록 개정과정을 통하여 법제화 되어 왔으나 앞으로의 생성형 AI시대를 위해서는 실시간으로 개인정보를 보호할 수 있는 또 다른 기술적인 노력[82]이 필요할 것으로 보인다.

---

81) L. Eliot. 각주 7).

82) 기존의 개인정보보호법은 제17조 4항에서 이미 암호화 등의 안전성 확보에 필요한 조치를 하였는지 여부를 고려하여 대통령령으로 정하는 바에 따라 정보주체의 동의 없이 개인정보를 제공할 수 있도록 하였다. 그렇다면 의료정보의 경우, 특별히 정기적인 보안 검사, 보다 엄격한 사용자 인증절차 등을 통해 특정 목적에 적합한 데이터만 수집할 수 있도록 따로이 규정하는 방법을 고려해 볼 수도 있을 것이다. 인공지능으로 인한 정보에 대한 윤리적 지침에 대한 견해로 J. Shaw. (2022). Emerging Paradigms for Ethical Review of Research Using Artificial Intelligence. American Journal of Bioethics Vol.22 Issue5.; 42-44 참조.

〔 참 고 문 헌 〕

<단행본>

고학수·김병필·구본효 외 4명,『인공지능 시대의 개인정보 보호법』, 서울: 박영사, 2022.

김윤명,『블랙박스를 열기 위한 인공지능법』, 서울: 박영사, 2022.

노갑영,『의료법과 의료분쟁의 이해』, 서울: 마인드탭, 2018.

이상돈·김나경,『의료법 강의』, 서울: 법문사, 2017.

송덕수,『신민법강의』, 서울: 박영사, 2022.

지원림,『민법강의』, 서울: 홍문사, 2022.

최윤섭,『의료 인공지능』, 서울: 클라우드 나인, 2018.

Isaac Kohane, Carey Goldberg, Peter Lee,『The AI revolution in Medicine: GPT-4 and Beyond』: Pearson Education, Inc. 2023.

<학술지>

김윤명, "생성형 인공지능(AI) 모델의 법률문제",『정보법학』 제27권 1호, 2023.

김병필, "대규모 언어모형 인공지능의 법적 쟁점",『정보법학』 제26권 1호, 2022.

김성원·최상민·이수원, "텍스트·데이터 마이닝 과정의 저작물 이용 면책 규정 신설안에 대한 소고-적대적 생성신경망에의 적용-",『지식재산연구』 제18권 1호, 2023.

김영두, "인공지능과 자유의지",『법학연구』, 제30권 1호. 2020.

김효연, "빅데이터와 정치(참여)에서의 세대 간 대표의 불균형 -정보격차와 디지털 리터러시를 중심으로-",『법학논고』, 제76집, 2022.

박노형, "개정된 개인정보보호법상 개인정보의 국외이전에 관한 규정의 분석: GDPR을 참조하여",『고려법학』 제109호, 2023.

박혜진, "의료 인공지능의 활용을 둘러싼 법적 과제: 규제의 진화 및 책임의 배분을 중심으로",『비교사법』 제29권 4호, 2022.

송미옥·조용진, "의료로봇의 현재와 미래: 수술로봇을 중심으로",『디지털융복합연구』 제19권 4호, 2021.

양종모, “인공지능의 위험의 특성과 법적 규제방안”, 『홍익법학』 제17권 4호, 2016.

엄주희·김소윤, “인공지능 의료와 법제”, 『한국의료법학회지』 제28권 2호, 2020.

유주선, “인공지능 의료행위와 법적 책임에 관한 연구”, 『비교사법』 제27권 4호, 2020.

유지혜, “인공지능 시스템에서 생성된 창작의 특허법상 보호에 관한 연구”, 『지식재산연구』 제18권 1호, 2023.

윤나리, “성명권의 내용 및 인격권에 기한 금지청구권의 인정요건에 관한 고찰-‘대법원 2022.11.17. 선고 2018다249995 판결’을 중심으로-”, 『법조』 제72권 4호, 2023.

윤용석, “위험책임분야의 민법개정의안서에 관한 일고”, 『법학연구』 제41권 1호, 부산대학교 법학연구소, 2000.

이계일, “독일의 법해석론에 대한 연구”, 『연세법학』 제38권, 연세법학회, 2021.

이동진, “일반적으로 접근 가능한 개인정보의 처리와 이익형량”, 『정보법학』 제24권 2호, 2020.

이부하, “일반적 인격권에 대한 헌법적 고찰-독일 연방헌법재판소 판례 고찰을 통한 일반적 인격권의 구체화-”, 『미디어와 인격권』 제8권 3호. 2022.

이수경, “산업용 AI로봇의 제조물 책임법 적용을 위한 소고”, 『정보법학』 제26권 1호, 2022.

이숙연, “인공지능 관련 규범 수립의 국내외 현황과 과제”, 『법조』 제72권 1호, 2023.

이해원, “개정 개인정보 보호법의 민사법적 쟁점”, 『법조』 제72권 3호, 2023.

정원준, “인공지능 창작과 저작권법의 딜레마”, 『고려법학』 제95권, 2019.

정채연, “의료 인공지능의 법적 수용을 위한 시론적 연구-쟁점과 과제-”, 『법학논총』 제45권 3호, 2021.

차상육, “인공지능(AI)과 지적재산권의 새로운 쟁점-저작권법을 중심으로-”, 『법조』 제66권 3호, 2017.

_____, “인공지능(AI) 관련 특허법상 쟁점에 관한 연구”, 『법학논고』 제80집, 2023.

A. Bernstein. “(Almost) No Bad Drugs: Near-Total Products Liability Immunity

for Pharmaceuticals Explained". 『Wash Lee Law Rev』. 77. 2020.

A. Holzinger, K. Keiblinger, P. Holub, et al. "AI for life: Trends in artificial intelligence for biotechnology". 『New Biotechnology』. 74. 2023.

A. Nashwan, A. Abujaber, H. Choudry. "Embracing the future of physician-patient communication: GPT-4 in gastroenterology". 『Gastroenterology and Hepatology』. 1(3). 2023.

B. Buchner. "Artificial intelligence as a challenge for the law: the example of "Doctor Algorithm"". 『Int Cybersecur Law Revw』. 3. 2022.

C.E. Haupt, M. Marks. "AI-Generated Medical Advice-GPT and Beyond". 『JAMA』. 3329. 2023.

D. Schwarcz, J.H. Choi. "AI tools for lawyers: a practical guide". 2023.

D.A. Drossman, J. Ruddy. "Improving patient-provider relationships to improve health care". 『J Clin Gastroenterol Hepatol』. 18(7). 2020.

H.R. Sullivan, S.J. Schweikart. "Are Current Tort Liability Doctrines Adequate for Addressing Injury Caused by AI?". 『AMA J Ethics』. 160. 2019.

J. Shaw. "Emerging Paradigms for Ethical Review of Research Using Artificial Intelligence". 『Am J Bioeth』. 22. 2022.

M. Hull. "Medical language proficiency: A discussion of interprofessional language competencies and potential for patient risk". 『Int J Nurs Stud』. 54. 2016.

M. Tsui. "Pharmaceutical product liability and the Australian Consumer Law: To wards a principled approach". 『Competition and Consumer Law Journal』. 23(2). 2015.

M.A. Stewart. "Effective physician-patient communication and health outcomes: a review". 『CMAJ』. 152. 1955.

M.D. Silva, T. Horsley, D. Singh et al. "Legal concerns in health-related artificial intelligence: a scoping review protocol". 『Systematic Reviews』.11. 2022.

N. Naik, B.M.Z. Hameed, D.K. Shetty, et al. "Legal and ethical consideration in artificial intelligence in healthcare: who takes responsibility?". 『Fron

Surg』. 14. 2022.

N. Sauder, A.K. Emara, P.J. Rullán, et al. “Hip and Knee are the most litigated Orthopaedic Cases: A nationwide 5-year analysis of medical malpractice claims”. 『J Arthroplasty』. 38. 2022.

P. Hacker. “Die regulierung von ChatGPT et al.- ein europäisches Trauersiel”. 『GRUR』. 289. 2023.

R.A. Khan, M. Jawaid, A.R. Khan, et al. “ChatGPT-Reshaping medical education and clinical management”. 『Pak J Med Sci』. 39. 2023.

S. O’Sullivan, N. Nevejans, C. Allen, et al. “Legal, regulatory, and ethical frameworks fo r development of standards in artificial intelligence (AI) and autonomous robotic surgery”. 『Int J Med Robot』. 15. 2019.

S. Vahdat, L. Hamzehgardeshi, S. Hessam, et al. “Patient involvement in health care decision making: a review”. 『Iran Red Crescent Med J』. 16. 2014.

S.D. Brown. “The Ethics of Medical AI and the Physician-Patient Relationship”. 『Camb Q Healthc Ethics』. 29. 2020.

S.S. Biswas. “Role of Chat GPT in Public Health”. 『Ann Biomed Eng』. 51. 2023.

T.B. Arif, U. Munaf, I. Ul-Haque. “The future of medical education and research: Is ChatGPT a blessing or blight in disguise?”. 『Medical Education Online』. 28. 2023.

Y. Ma. “The potential application of ChatGPT in gastrointestinal pathology”. 『Gastroenterology & Endoscopy』. 1. 2023.

<온라인 자료>

보건복지부, 개인정보보호위원회, “보건의료데이터 활용 가이드라인”, 2022.

[국문초록]

생성형 AI의 의료적 활용과 개인정보보호

이수경 (법학박사, 고려대학교 연구교수)

생성형 AI의 활용은 교육계를 넘어서 이미 의료계에서도 의료 기기에 임상 소프트웨어 등의 도입 등으로 연구되고 있다. 생성형 AI는 대규모 대화형 언어모델을 활용하여 방대한 데이터를 이해하고 자료를 선별하는 시간과 에너지를 줄여주면서 사용자와 끊임없는 대화를 통한 정보의 전달이 가능하다. 바로 이러한 점이 인류에게 생성형 AI가 혁신적인 기술의 등장으로 인정받고 있는 점이기도 하다. 그러나 반면 사용자에게 제공되는 컨텐츠의 정합성은 출처나 근거 없이 사용자에게 판단의 영역으로 맡겨지고 있다. 그러나 이 글에서는 생성형 AI를 활용함에 있어서 가장 직접적으로 발생할 수 있는 쟁점을 우선적으로 살펴보기로 한다. 따라서 이 글에서는 생성형 AI의 대표적인 프로그램인 Chat GPT의 발전과 이용자의 활용에 대비하여 특히 개인정보 보호의 쟁점에 대하여 논의하였다. 이를 위하여 먼저 생성형 AI의 기술적인 특성을 살펴본 뒤에 발생 가능한 민사적 쟁점 가운데에서도 개인정보 보호에 관한 문제를 우선적으로 살펴보았다. 생성형 AI는 그 자체로서 학습 데이터의 편향이나 출처 없는 결과값의 제공 등 여러 문제점이 제기되고 있으나, 이러한 문제점은 윤리적 문제를 내포하는 것으로 당장 임상 소프트웨어로서 의료기기에서 활용될 경우 개인정보 보호법제와 보건의료데이터의 활용 가이드로 환자 혹은 이용자의 개인정보를 보호할 수 있을 것인가에 대한 의문에 대한 논의가 시급하다고 판단되었다. 우리나라의 개인정보 보호법제는 특히 보건의료데이터의 활용에서 특정 개인의 개인정보를 가명처리하고 비식별조치를 취하는 데에 적절한 프로세스를 갖추고 있는 것으로 보이나, 생성형 AI이 소프트웨어로서 의료기기에 적용되었을 경우에도 이 법제로서 개인정보 보호의 목적을 이루기에는 어려운 점이 있다. 임상 소프트웨어에서 활용될 생성형 AI의 기능을 대비하기 위해서는 생성형 AI에 걸맞는 개인정보 보호의 법제가 필요할 것으로 보인다.

주제어 : 생성형인공지능, 개인정보 보호법, 보건의료데이터, 인공지능, 비식별조치

A Study on the Medical Application and Personal Information Protection of Generative AI

Lee, Sookyoung

Ph.D. in Law, Research Professor, Korea University

=ABSTRACT=

The utilization of generative AI in the medical field is also being rapidly researched. Access to vast data sets reduces the time and energy spent in selecting information. However, as the effort put into content creation decreases, there is a greater likelihood of associated issues arising. For example, with generative AI, users must discern the accuracy of results themselves, as these AIs learn from data within a set period and generate outcomes. While the answers may appear plausible, their sources are often unclear, making it challenging to determine their veracity. Additionally, the possibility of presenting results from a biased or distorted perspective cannot be discounted at present on ethical grounds.

Despite these concerns, the field of generative AI is continually advancing, with an increasing number of users leveraging it in various sectors, including biomedical and life sciences. This raises important legal considerations regarding who bears responsibility and to what extent for any damages caused by these high-performance AI algorithms.

A general overview of issues with generative AI includes those discussed above, but another perspective arises from its fundamental nature as a large-scale language model (‘LLM’) AI. There is a civil law concern regarding “the memorization of training data within artificial neural networks and its subsequent reproduction”. Medical data, by nature, often reflects personal characteristics of patients, potentially leading to issues such as the regeneration of personal information. The extensive application of generative AI in scenarios beyond traditional AI brings forth the possibility of legal challenges that cannot be ignored.

Upon examining the technical characteristics of generative AI and focusing on legal issues, especially concerning the protection of personal information, it's evident that current laws regarding personal information protection, particularly in the context of health and medical data utilization, are inadequate. These laws provide processes for anonymizing and de-identification, specific personal information but fall short when generative AI is applied as software in medical devices. To address the functionalities of generative AI in clinical software, a reevaluation and adjustment of existing laws for the protection of personal information are imperative.

Keyword : Generative AI, Personal information protection act, Healthcare date, Artificial Intelligence, De-identification.