민변디지털정보위원회, 정보인권연구소, 진보네트워크센터,
보건의료단체연합, 참여연대

수 신	각 언론사 정치부·사회부
발 신	참여연대 공익법센터 (담당 : 이지은 간사 02-723-0666) 진보네트워크센터 (담당 : 오병일 대표 02-701-4551)
제 목	[보도자료] < “세계최초” 인공지능법안, 세계의 걱정거리가 되려는가?> 기자설명회 개최
날 짜	2023. 3. 22. (붙임자료 포함, 별첨자료 제외 총 13쪽)

보 도 자 료

<과방위 소위 통과 “세계최초” 인공지능법안, 세계의 걱정거리가 되려는가?> 기자설명회 개최

“선허용 · 후규제”는 안전과 생명 포기하겠다는 것
고위험 인공지능 정의 자의적이고 구체적 위험방지 대책도 전무
기본권 보호 헌법가치와 충돌하여 전면 재검토 필요

1. 오늘(3/22) 민변 디지털정보위원회, 정보인권연구소, 진보네트워크센터, 보건의료단체연합, 참여연대는 지난 2/14 인공지능과 관련한 법안 7개가 병합되어 국회 과학기술정보방송통신위원회 법안심사소위를 통과한 인공지능법안에 반대하며 이 법안의 구체적인 문제점 등에 대해 기자설명회를 개최하였습니다.
2. 단체들은 과방위 소위 통과 인공지능법안은 인공지능 규율을 위한 기본법임을 표명하지만, 안전과 인권에 미치는 인공지능의 위험을 규제할 수 있는 실질적인 내용은 없다고 평가했습니다. 또한 국민의 안전과 인권을 보호하는 법제를 준비하기는커녕 산업 육성만을 위해 ‘우선허용 사후규제’라는 원칙을 도입하여 정당한 규제의 도입을 방해할 가능성이 크다고 지적하였습니다. 또한 정보인권 보호, 소비자 보호, 차별 금지 등에 관여하는 규제기관이 인공지능법안의 관할기관을 맡는 유럽연합이나 미국 등과 달리 산업 육성만을 위해 경도된 입장으로 일관하던 과기부가 관할하도록 하여 세계적으로도 드문 입법례라고 설명하였습니다.
3. 오늘 기자설명회는 김태일 참여연대 권력감시1팀장의 사회로 진행되었으며, 희우 진보네트워크센터 활동가가 <인공지능법안의 제정 경과 및 주요 쟁점에 대한 개요>를, 장여경 정보인권연구소 상임이사가 <고위험인공지능 분류 및 규제 관련 유럽연합과 미국

등 해외 입법례와 과방위 통과법안의 근본적 차이>를, 허진민 참여연대 공익법센터 소장이 <과방위 통과법안의 독소조항 등 문제점>을, 김병욱 민변 디지털정보위원회 변호사가 <과방위 통과법안이 타법과 충돌할 가능성, 타 규제기관의 작용을 방해하는 지점>을, 그리고 마지막으로 보건의료단체연합 전진한 정책국장이 <보건 및 의료 관련 인공지능 적용 사례를 통해 본 위험성과 특별한 규제 필요성>을 발표하였습니다. 끝.

▣붙임자료

1. 희우 진보네트워크센터 활동가 발표자료
2. 허진민 참여연대 공익법센터 소장 발표자료
3. 김병욱 민주사회를 위한 변호사모임 디지털정보위원회 위원 발표자료
4. 전진한 보건의료단체연합 정책국장 발표자료

▣별첨자료1. 장여경 정보인권연구소 상임이사 발표자료

- 고위험인공지능 분류 및 규제 관련 유럽연합과 미국 등 해외 입법례와 과방위 통과법안의 근본적 차이 - 장여경 정보인권연구소 상임이사 (http://bit.ly/3FDn1ne)

▣붙임1. 희우 진보네트워크센터 활동가 발표자료

인공지능법안의 제정 경과 및 주요 쟁점에 대한 개요

희우 진보네트워크센터 활동가

지난 2월 14일 과학기술정보방송통신위원회 심사소위를 통과한 인공지능산업 육성에 관한 법안은 지금까지 계류되어 있던 7개 법안을 가장 최근에 발의된 국민의힘 윤두현 의원안을 중심으로 병합한 결과물입니다. 심사소위를 통과해 법사위와 본회의 심의만을 남겨두었기 때문에 언론에서 이 법안에 대해 '8부 능선을 넘었다'고 표현할 만큼 제정이 목전입니다.

하지만 인공지능이 어떤 영향력을 가지게 될 지, 어떤 범위까지 뻗어나갈 지 아직 알 수 없는 시점인데도 불구하고 이 법안은 인공지능 산업에 대한 국민 안전 및 인권 보장 규제를 완화하며, 대부분의 규범을 과학기술정보통신부가 담당한 채 '선허용, 후규제'한다는 일견 무책임한 내용을 담고 있습니다.

앞서 말씀드렸듯이 인공지능은 현재 광범위한 영향을 끼치고 있습니다. 학교나 가정에서 사용되는 인공지능 스피커, 의료진단에 사용되는 인공지능, 보험이나 대출 등 금융서비스에서 사용되는 알고리즘, 검색이나 배달앱 등 플랫폼에서 사용되는 알고리즘, 출입에 사용되는 얼굴인식 알고리즘 등, 어떤 인공지능은 생계나 안전, 인권에 이미 영향을 미치고 있는 상황입니다. 이러한 영향력은 앞으로 커지면 커졌지 결코 줄어들 수 없으며, 특히 앞으로 사람에게 영향을 미치는 의사결정이 인공지능에 기반해 이루어지는 사례가 더욱 많아질 것으로 예상됩니다.

한국에서도 이미 인공지능과 관련된 논란이 몇 차례 있었습니다. 먼저 2020년 12월에 런칭된 챗봇 이루다 사건입니다. 대화형 챗봇 '이루다'를 개발하고 운영하는 스캐터랩은 2013년 텍스트앳, 2016년 연애의 과학 등의 어플리케이션을 운영하며 카카오톡 등 메신저의 대화 내용을 수집해왔고 이를 자사 다른 제품인 대화형 챗봇 '이루다'의 학습용 데이터로 이용했습니다. 하지만 이용자들에게 그 사실을 제대로 고지하지 않은 채 수집 동의를 받고 데이터와 알고리즘을 깃허브에 업로드해 개인정보를 노출하는 등 개인정보보호법을 위반했으며, 학습한 데이터를 토대로 답변하는 과정에서 성소수자에 대한 혐오 발언을 하는 등 윤리적 문제점도 지니고 있었습니다. 제조사인 스캐터랩은 1억 330만원의 과징금과 과태료를 부과받았으며 챗봇 이루다는 런칭 3주만에 서비스가 중단되었습니다.

또한 공공기관의 인공지능 채용 프로그램을 둘러싼 논란도 있었습니다. 시민사회단체들은 2020년에 공공기관에서 사용하는 인공지능 채용도구에 대한 실태를 파악하기 위해 13곳의 공공기관에 정보공개를 청구했습니다. 이는 공공기관이 민간업체의 인공지능 채용 도구를 도입하면서 인공지능의 차별성 및 편향성에 대해 사전에 검토를 했는지, 면접자들의 개인정보에 대해 적절한 보호를 하고

있는지, 공공기관으로서의 인공지능 운영 과정에 투명성과 책무성을 담보하고 있는지를 확인하기 위함이었습니다.
그러나 대부분의 기관들이 자료의 부존재, 영업비밀 보호 등의 이유로 자료를 공개하지 않았고, 단체들이 소송을 제기해 승소함으로써 두 기관의 일부 비공개 정보를 공개하게 되었는데, 기관이 보유하고 있지 않은 정보들은 당연히 공개될 수 없었습니다.
즉, 공공기관이 민간기업의 인공지능 도구를 도입하면서도 그 인공지능의 문제점과 성능에 대해 아무런 검토를 하지 않았으며, 공공기관으로서 민원에 답할 수 있는 적절한 자료들을 보유하고 있지 않았다는 점이 드러난 것입니다. 이 사례는 한국의 공공기관들이 인공지능 시스템을 도입하고 있음에도 불구하고, 공공기관의 책무성을 담보할 수 있도록 하는 아무런 제도적 장치가 존재하지 않는다는 점을 보여주고 있습니다.

또한 법무부와 과기부가 추진한 '인공지능 식별추적 시스템'도 논란이 되었습니다. 2021년 10월, 법무부와 과기부가 2019년부터 법무부가 출입국심사과정에서 수집, 보유하고 있는 내·외국인의 안면 데이터 약 1억 7천만 건을 정보주체의 동의없이 민간기업들에게 인공지능 학습 및 알고리즘 검증용으로 제공한 사실이 드러났습니다.
이 사건은 과기부가 공공기관이 보유하고 있는 개인정보를 제공하여 인공지능 기업들의 기술 개발을 위해서라는 점을 공공연하게 언급하고 있는 점, 그리고 경쟁적인 R&D를 통한 공모방식으로 추진되어 실제 위탁 업체로 선정되지 않을 업체에게도 민감한 개인정보가 제공되었다는 점이 논란의 대상이 되었습니다.
또한, 얼굴인식 기술에 대한 국가 감시 가능성이 세계적인 논란이 되고 있음에도 불구하고, 정부가 인공지능 규제를 위한 법제도 없는 상황에서 내·외국인의 얼굴인식 정보를 대량으로 제공했다는 점은 한국 정부의 인공지능 정책이 인권보다는 산업 육성을 우선하고 있음을 보여주고 있습니다.

위 사례의 문제들은 인권위나 개보위가 개입해 문제를 조율하는 역할을 했는데, 산업 육성을 우선시하는 과기부가 해당 법안을 통해 인공지능 관련 최상위 기관이 된다면, 위 사례와 같은 문제가 발생하거나 인공지능 개발 및 활용 과정에서 다른 문제가 발생했을 시 제대로 해결할 수 있을지 의문입니다.

또한 2019년 12월 17일, 과기부를 비롯한 전 부처 공동으로 <인공지능(AI) 국가전략>을 발표했는데, 9대 전략 중 하나인 전략 (3) 과감한 규제혁신 에서 '선허용-후규제'를 기본방향으로 제시하고 있었습니다.

이어 2020년 12월 23일, 과기부와 정보통신정책연구원은 <인공지능(AI) 윤리기준>을 발표했는데, 이 보도자료에서 AI 윤리기준은 "구속력 있는 '법'이나 '지침'이 아닌 도덕적 규범이자 자율규범"임을 밝히고 있습니다.

이어 과기부는 2021년 5월 13일, <신뢰할 수 있는 인공지능 실현전략>을 발표한 바 있는데, 이 역시 '민간 자율'적으로 신뢰성을 확보할 수 있도록 지원체계를 구축하는 것에 중점을 두고 있습니다.

이처럼 과기부는 인공지능의 위험성에 대한 통제보다는 민간 자율과 윤리를 통한 규율, 선허용 후규제 도입이라는 기조를 유지하며 산업육성에만 치중해 왔습니다.

이 때문에 시민사회는 과기부가 인공지능 규율의 주무부처가 되는 것에 대해 우려를 표하며, 같은 해 5월 인공지능 입법에 있어 인공지능 제품과 시스템을 국가적으로 감독하는 체계를 수립하여야 하고, 인공지능을 감독하는 역할은 산업부처나 기술부처가 아니라 공정위, 인권위, 개보위가 수행해야 마땅하다고 지적한 바 있습니다.

반면 정부와 다르게, 감독을 수행할 수 있는 국가인권위 등 독립적 기구들은 인공지능의 위험성에 대해 통제의 필요성을 제안해 왔습니다.

인권 관점의 인공지능 규율을 위한 가이드라인은 개보위에서 먼저 나왔는데, 2021년 5월 31일, 개보위는 인공지능(AI) 서비스의 개발과 운영 시 발생할 수 있는 개인정보 침해를 예방하기 위한 안내서로서 <AI 개인정보보호 자율점검표(개발자·운영자용)>를 공개했습니다.

이어 2022년 5월 11일, 국가인권위가 인공지능 개발 및 활용 과정에서 발생할 수 있는 인권침해와 차별을 방지하기 위한 <인공지능 개발과 활용에 관한 인권 가이드라인>을 발표했습니다. 인권위는 이 가이드라인에 기초해 인공지능 관련 정책이 수립, 이행되고, 관계 법령이 제/개정되도록 관련 부처들을 유기적으로 조정하고 통할할 것을 국무총리에 권고했습니다. 또한 과기부, 개보위, 방통위, 공정위, 금융위에 이 가이드라인에 기초하여 관련 정책을 수립하고 법령을 제개정하며, 공공기관 및 민간기업도 이를 준수할 수 있도록 감독할 것을 권고한 바 있습니다.

해당 권고에 대하여 국무총리와 과기정통부 장관, 개보위 위원장, 방통위 위원장, 공정위 위원장, 금융위 위원장 등은 해당 업무와 관련한 인권위의 권고 취지에 공감하며, 관련 정책과 사업 및 제도 개선에〈가이드라인〉의 내용을 반영하겠다고 회신한 바 있습니다.

그러나 해당 인공지능 법안이 인권위의 가이드라인을 수용하고 있는지는 의문입니다.

또한, 인권위는 2022년에 인공지능에 대한 인권적 통제방안으로서 인권영향평가 제도에 대하여 검토하고 그 실현 방안을 연구한 <인공지능 인권영향평가 도입방안> 연구를 진행하여, 2023년 초에 공개했습니다.

그리고 2023년 1월 25일, 인권위는 얼굴인식 기술이 인권을 침해하지 않도록 입법으로 보호하고 인권영향평가를 실시할 것을 정부와 국회에 권고했습니다. 특히 불특정 다수를 대상으로 공공장소에서 '실시간 원격 얼굴인식'을 사용하는 것은 기본권 침해 위험이 매우 크다고 지적하면서 원칙적 금지를 요구했는데, 과연 이번 인공지능 법안이 이러한 국가인권위 권고를 이행할 수 있는 것인지 의문입니다.

한국의 시민사회단체 역시 인공지능 관련 정책에 적극 목소리를 내고 있습니다. 2021년 5월 24일, 120개 시민사회단체는 함께 <인권과 안전, 민주주의가 보장되는 인공지능 정책 요구 시민사회 선언문>을 발표한 바 있습니다.

현재 주요 국가들은 인공지능의 위험을 효과적으로 규제하고, 이를 위한 독립적인 감독체계를 수립하며, 권리구제에 대한 내용을 구체화하는 기조로 인공지능이 안전과 인권에 미치는 위험을 규제하기 위한 입법을 추진해 왔습니다.

만일 국가적 수준에서 인공지능에 대한 기본법을 제정한다면, 다른 규제기관의 업무와 조화를 이루며 인공지능이 야기할 수 있는 위험으로부터 국민의 안전과 인권을 지키고, 국민의 피해를 구제하기 위한 장치를 충분히 마련해야 할 것입니다.

▣붙임2. 허진민 참여연대 공익법센터 소장 발표자료

과방위 통과법안의 독소조항 등 문제점

허진민 참여연대 공익법센터 소장

- 인공지능의 개념 자체가 정확하게 정의되어 있지 않지만 머신러닝을 기반으로 하여 인간의 사유, 추론을 대체할 수 있는 기술을 의미한다면 사회적 패러다임의 변화를 예견하는 것임.
- 이처럼 인간만이 수행하던 사유, 추론 등을 소프트웨어로 구현한 프로그램 등으로 수행이 가능하게 되므로 이러한 프로그램 등의 자동 과정과 결과에 대한 안전성과 신뢰성의 확보가 필연적이라 할 것임. 그리고 이러한 안전성과 신뢰성의 확보를 위한 기준은 사회적 합의에 따라 정해져야 하며 이에 다른 나라들도 인공지능기술의 안전성과 신뢰성을 담보하기 위한 논의가 지속되고 있음.
- 그러므로 인공지능기술의 정의 자체도 명확하지 않으며, 인공지능기술은 향후 인간이 수행해온 상당 부분의 역할을 대체할 수밖에 없음에도, 인공지능기술의 특성이 스스로 학습을 통한 결과 도출인 점을 고려하면 인공지능기술의 안정성과 신뢰성 확보가 전제되어야함에도 이에 대한 고려 없이 성급하게 인공지능기술 육성에만 치중한 과방위 통과법안은 부적절함.
- 우리 헌법의 최고 가치는 “인간의 존엄성” 보장이며 국가는 국민의 기본권을 보호할 의무를 가짐. 한편 헌법은 국가로 하여금 과학기술의 혁신을 통하여 국민경제의 발전에 노력할 의무를 부과하고 있음. 그러나 이는 국민의 기본권을 보호하고 국만경제의 발전에 부합해야 하는 한계를 가짐. 따라서 이번 법안이 이와 같은 헌법가치에 부합하나 살펴보아야 할 것임.

- 우선 첫째 과방위 통과법과 다른 법률과의 관계 법체계적 관점에서 현행 지능정보화기본법에서 인공지능을 포함한 다양한 신기술에 대한 육성 등에 대한 규정을 두고 있어 별도의 인공지능법이 필요한지 의문이며, 지능정보화기본법은 지능정보화사회외 관련한 기본법의 성격을 가짐에도 불구하고 과방위 통과법은 인공지능산업 육성을 위한 법임에도 안 제4조에서 기본법으로서의 지위를 부과하고 지능정보화기본법과의 관계에서 법체계적으로 조화될 수 없는 규정을 두고 있어 법안이 통과되면 지능정보화기본법을 따라야 할지 이법을 따라야 할지 혼란을 준다는 점에서 문제가 될 것임.

- 둘째, 산업육성이라는 입법목적을 달성하려면 인공지능기술의 안전성과 신뢰성의 담보없이는 산업의 발전이 불가능 함. 특히 안 11조는 인공지능기술이 사용될 영역에 대한 어떠한 제한도 없이 선허용 후규제 형태의 조항을 두고 있으며, “생명, 안전 공공복리를 현저히 저해할 우려”가 있을 때 규제한다고 하더라도 그 “현저히 저해”할 우려가 어떤 것인지에 대해 명확한 기준이 없음. 우선 진입하고 나중에 문제가 생기면 규제하겠다는 방식 자체는 오히려 산업육성의 목적과는 부합하지 않음. 국가가 안정성과 신뢰성의 기준을 제시하고 인공지능기술의 사용이 제한되는 영역을 명확하게 설정해주어야 함. 그렇지 않다면 인공지능기술의 개발이 이루어지고 상용화되는 과정에서 규제가 이루어진다면 사회적 혼란을 야기하고 경제적 손실만을 초래할 것임. 따라서 안 11조는 삭제되어야 함.
- 셋째, 안 제26조 및 안제27조는 인간의 생명, 신체의 안전 및 기본권에 중대한 영향을 미칠 우려가 있는 고위험영역과 관련하여 고위험영역인공지능의 해당 여부를 과기부장관에게 확인받을 수 있고, 이를 제공함에 있어 이용자에게 사전고지할 의무만을 부과하고 있음.
- 우선 고위험영역에서 인공지능기술이 사용될 수 있는지 여부 및 사용을 허가한다면 어떠한 조건을 부과할 것인지에 대하여 사회적 합의가 도출되지 않았으며, 외국의 경우 그러한 기준을 도출하는 과정에 있음에도 과방위 통과법은 고위험영역의 인공지능기술의 사용 여부 및 조건에 대한 어떠한 내용도 담고 있지 않으며, 심지어 고위험영역인공지능의 해당 여부에 대한 확인 조차 개발/판매자가 임의적으로 선택할 수 있도록 하여 국민의 생명, 안전 기본권 침해에 대하여 어떠한 보호 조치가 없음을 보여주는 것으로 폐기되어야 함.

▣붙임3. 김병욱 민주사회를 위한 변호사모임 디지털정보위원회 위원 발표자료

인공지능법안이 타 규제기관의 작용을 방해할 가능성과 그 지점

김병욱 변호사(민변 디정위)

- 법안은 과학기술정보통신부에 인공지능 관련 규율 및 정책 일반에 대한 폭넓은 권한을 부여하면서, 국가 및 지방자치단체가 인공지능기술, 인공지능제품 또는 인공지능서비스와 관련한 법령 및 제도를 도입할 때 '우선허용 사후규제 원칙'을 따르도록 정하고 있음(법안 제11조 제2항).

- 법안 제11조 제1항에 '우선허용 사후규제 원칙'의 예외가 규정되어 있으나, 매우 추상적인 수준에 머물러 있음(국민의 안전, 생명, 권익에 위해가 되거나 공공의 안전보장, 질서유지 및 복리증진을 현저히 저해할 우려). 예외 규정의 의미가 모호하면서 엄격하기 때문에 인공지능에 대한 사전적인 규율 내지 규제 가능성을 무력화시키는 조항으로 기능할 우려가 상당함. 인공지능기술과 피해 사이에 구체적인 인과성이 인정되기 이전에 위해가 되거나 현저히 저해할 우려가 인정되기 어렵기 때문에 사실상 사전적인 개입이나 규제를 배제하는 결과가 될 것임.

- 그러나 인공지능기술은 다양한 분야에서 매우 광범위하게 활용되고 있고, 다양한 기본권을 제약하거나 침해할 가능성을 지니고 있음. 그 위험성을 사전적으로 감시, 감독하고 위험을 관리하여 침해를 최소화하여야 마땅함. 다양한 분야에서 활용되는 인공지능에 대해 각 분야별 소관 부처가 관련 시책을 마련할 수 있고, 이는 사전적인 규제를 포함하는 것이나, 위 법안에 의하여 이러한 조치는 제약을 받거나 무력화될 수밖에 없음.

- 예를 들어 인공지능 기술을 채용 분야에 활용하는 것과 관련하여, 개인정보자기 결정권 보장의 관점에서, 인공지능 프로그램이 고도화되는 과정에서 개인정보가 적법하게 수집되고, 활용되었는지, 프로그램 활용 과정에서 개인정보의 수집이 적법한지, 이후 보관 및 처리 등의 과정이 적법한지 등에 대해서 개인정보보호위원회가 관련 법제 정비나 가이드라인 마련 등을 통해서 사전에 개입할 수 있고, 고용노동부가 채용절차의 공정성의 관점에서 인공지능 프로그램이 직무에 필요하지 않은 불필요한 개인정보를 수집하지 않는지, 인공지능프로그램의 의사결정에 따라 채용 여부가 결정되는 경우 불복하거나, 구제가능성을 보장할 수 있는지 등에 대하여 사전에 개입하여 프로그램의 도입을 재검토하거나 보완할 수 있을 것이나, 위 법안에 의하면 이러한 시도가 제약되거나 무력화될 가능성이 있음.

- 인공지능 기술을 탑재한 제품이 인간에게 위해를 가하는 등 신체의 안전을 위협할 가능성도 배제할 수 없는데, 이 경우 제품 안전 관련 부처인 산업통상자원부 등이 사전에 인공지능기술의 위험성을 관리, 감독하거나 특정한 인증을 거치기 전에는 제품을 출시할 수 없도록 하는 등의 법제 등 규제를 마련할 수 있을 것이나, 위 조항에 따르면 이러한 시도는 제약되거나 무력화될 가능성이 있음.

▣붙임4. 전진한 보건의료단체연합 정책국장 발표자료

보건 및 의료 관련 인공지능 적용 사례를 통해 본 위험성과 특별한 규제 필요성

전진한 보건의료단체연합 정책국장

이 법안은 의료기기에도, 보건의료 전반에 적용하는 인공지능에도 우선허용 사후규제를 적용합니다.

제대로 검증되지 않은 의료 인공지능이 도입됐을 때 발생할 수 있는 폐해에 대해서 예를 들어 말씀드리고자 합니다.

의료 인공지능으로 가장 잘 알려진 건 IBM이 개발한 '왓슨 포 옹콜로지'입니다. 왓슨은, 의사가 암환자 데이터를 입력하면 치료방법을 제시하는 프로그램이었습니다. IBM은 이 프로그램을 '암 치료의 혁명'이라고 홍보했습니다.

문제는 이 기술이 충분히 검증되지 않았다는 것입니다. 왓슨은 안전하지 않고 부정확한 치료법을 추천했습니다. 폐암의 경우 정확도가 18%, 위암과 유방암의 정확도도 40%대에 불과해서 어떤 의사들은 이 프로그램을 "쓰레기"라고 하기도 했습니다.

그럼에도 불구하고 많은 병원들이 '왓슨'을 도입했습니다. 그 이유는 과장된 홍보로 암환자를 유인할 수 있고 인공지능을 쓴다는 이유로 추가 비용을 청구할 수 있었기 때문입니다. 한국의 대학병원들도 너도나도 이 프로그램을 도입해서 환자를 끌어들였습니다. 길병원, 부산대병원, 건양대병원, 대구가톨릭대, 계명대, 조선대, 화순전남대병원 등이 이 '쓰레기'를 도입해서 환자를 유인했습니다.

이것이 보여주는 바는, 규제되지 않은 인공지능은 최악의 경우에 환자의 건강과 생명을 위협할 수 있고, 가장 운이 좋은 경우에도 국민들이 불필요한 비용을 지출하게 만든다는 것입니다. 기업과 병원의 수익창출을 위해서. 결국 최근에 IBM이 왓슨을 헐값에 매각하면서 이 프로젝트는 실패로 끝났습니다. 하지만 이미 환자들에게 부정적 영향을 미친 뒤입니다. 심지어 지금도 여전히 많은 병원 홈페이지에 왓슨을 이용해서 암치료를 한다는 홍보 게시물이 올라가 있습니다.

영국에서는 '바빌론'이라는 이름의 AI 의료 챗봇 서비스가 도입이 됐습니다. 바빌론은 인공지능 챗봇으로 환자를 미리 걸러 치료가 필요한 환자만 치료를 받게 해서 국가 의료비용을 절감시키겠다고 약속했습니다. 영국 정부는 이 인공지능이 효과가 있다는 근거도 없이 승인했습니다. 실제로 바빌론 챗봇은 불충분하거나 명백하게 잘못된 정보를 환자에게 전달했습니다. 그래서 질병이 있는 사람들의 치료가 지연되거나 차단됐습니다. 그래서 학계에서는 검증되지 않은 인공지능을 허용하는 것은 마치

신약을 제대로 테스트하지 않고 환자에게 투여하는 것과 다를 바가 없는 위험천만한 행위라고 평가합니다.

바빌론도 지난 해 말 영국정부로부터 계약해지되었습니다. 하지만 뒤늦은 결정은 피해를 되돌리지 못합니다.

지금까지 말씀드린 것은 인공지능이 여타 의료기술들과 마찬가지로 충분히 검증하지 않으면 환자 생명과 건강을 침해할 수 있다는 것이었습니다.

그런데 여기서 그치지 않습니다. 인공지능은 여타 의료기술들보다 더 위험할 수 있고 윤리적 문제를 일으킬 때문에 더 충분한 기술적, 사회적 검토가 필요합니다.

첫째, 의료서비스 제공자 개인의 오류와는 달리, 인공지능 알고리즘에 오류가 있으면 그것은 단기간에 수천 수만명의 사람들에게 돌이킬 수 없는 피해를 줄 수 있습니다.

예를 들어 영국 코로나19 애플리케이션은 감염자와 밀접 접촉할 경우에 자가격리를 지시하도록 설계됐는데 기능에 문제가 있었습니다. 그래서 실제 위험보다 5배는 더 오래 전염성이 있는 환자 곁에 머물게 했습니다. 1900만명이 앱을 다운로드 했는데 엄청나게 적은 수만 격리하라는 지시를 받아서 자신과 가족들을 감염에 노출시켰습니다.

아까 말씀드린 바빌론 챗봇도 마찬가지로 광범한 인구에 영향을 준 사건입니다.

그래서 어떤 연구는 디지털 기술의 오류와 결함으로 영국에서 연간 2천명이 사망할 수 있다면서 이것이 '눈에 띄지 않는 살인자'라고 했습니다.

둘째, 인공지능 기술은 불투명한 경향이 있고 '블랙박스' 알고리즘을 사용할 수 있기 때문에 의사결정을 내리는 과정과 기준을 알기 어려운 경우가 많습니다.

그래서 이런 오류는 교정되기가 더 어려울 수 있습니다.

또 피해가 발생할 경우 책임을 누가 질 것인가 하는 문제가 발생합니다.

이런 불투명한 기술을 누군가가 비윤리적으로 설계하면서 이익을 창출할 수 있습니다.

예를 들면 미국에서 인공지능은 의료보장을 줄이는 데 활용되고 있습니다. 실제 일어난 일에 따르면, 골절로 입원한 노인 환자에 대해서 인공지능은 17일 후 퇴원할 수 있다고 예상을 했습니다. 17일째 되는 날에 보험사는 알고리즘에 따라서 치료비 지급을 중단했습니다. 하지만 환자의 통증은 극에 달했을 때였습니다. 이 결정을 뒤집는 법원 결정이 나는 데 1년이 걸렸습니다. 미국에서 이런 방식의 보험금 지급거부가 새로운 방식으로 떠오르고 있습니다. 3개월 내 사망할 수 있는 환자 보험금 지급을 거절해서 최대 2~3년이 걸리는 이의신청 절차를 밟게 만들고 있습니다.

셋째, 차별과 편견을 강화할 수 있습니다.

바빌롯 챗봇에서는 성별 편향성도 발견됐습니다. 예를들면 흉통과 메스꺼움을 호소하는 여성의 경우에는 이 인공지능이 우울증이나 공황발작 가능성을 제시했고, 비슷한 증상의 남성에게는 심각한 심장 문제의 가능성을 언급하면서 응급실 방문을 권장했습니다.

미국의 비슷한 프로그램에서도 이런 문제가 발생했습니다. 비슷한 수준의 의료서비스가 필요한 경우에 대체로 백인 환자가 흑인 환자보다 더 아프다고 판단하고 흑인에게 더 적은 의료를 제공했습니다.

이것은 '의도적 설계'나 '기술적 오류'가 아니라 사회의 편견이 인공지능 데이터에 반영된 것일 수 있습니다.

따라서 AI기술은 처음부터 '윤리적 설계'를 제대로 적용해야 합니다.

넷째, 설령 기술적 문제가 전혀 없다고 해도 윤리적, 사회적 문제를 낳을 수 있습니다.

예컨대 어떤 개인이 당뇨나 HIV감염 가능성이 있는지 예측하는 인공지능 기술을 생각해볼 수 있습니다. 그 경우에 이런 기술은 특정 개인이나 커뮤니티에 혜택을 줄수도 있지만 문제의 책임을 당사자들에 돌리면서 불필요한 낙인을 찍을 수 있고, 영리 목적의 건강관리서비스로부터 공격적 마케팅에 노출시킬 수 있습니다.

이런 건강 예측모델을 고용주나 보험사가 활용하면 개인의 자율성을 침해하고 일상생활을 통제할 수도 있습니다.

마지막으로, 인공지능을 적용하면 개인정보 유출 문제가 더욱 빈번할 수 있고, 인공지능 알고리즘에 대한 해킹공격은 막대한 피해를 입힐 수 있습니다.

이런 인공지능 기술을 선진입-후평가 하는 것은 말도 안되는 것입니다. 그 결과는 엄청난 재앙일 수 있습니다. 오히려 인공지능은 기존기술보다 더 엄격한 안전성과 효과에 대한 평가, 윤리적 사회적 검토, 개인정보 보호를 위한 엄격한 기준 등이 필요합니다.

이 법이 의료를 포함한 몇몇 분야 인공지능은 고위험 기술이라고 분류했습니다. 하지만 거의 아무런 제한을 두지 않고 있습니다. 그것은 이 법이 위험하다는 걸 알면서도 국민의 안전을 포기하겠다는 것입니다.

저는 보건의료 영역에 대해서만 말했지만 사회 각 부문에서 검증되지 않은 인공지능의 부정적 영향이 막대할 수 있습니다.

인공지능 검증을 생략하는 이런 법은 반드시 철회되어야 하고 인공지능을 가른 기술들보다 더 엄격히 규제하는 입법이 필요합니다. 끝.