설명의무 관점에서 본 알고리즘 공개와 영업비밀

김윤명
경희대학교

오래전 이야기

왜, 알고리즘 공정성 논의인가?

- 알고리즘을 악용하거나 오용함으로써, 새로운 차별을 만들어 내거나 기존의 차별을 고착화시킨다는 문제가 제기되고 있다.
- 알고리즘의 영향을 받을 대중이 관심을 가져야할 사안으로, 치밀하게 은연 중에 이루어지는 경우라면, 인간의 능력으로는 대응방안을 찾기가 어려울 수 있다.
- (참고) 현행법상 공정성 정의는 없으나, 목적의 공정성과 절차의 정당성을 확보해야하는 것!

학습데이터의 내재적 한계

- “기계학습은 데이터 분석 능력을 크게 향상시킴으로써 사회에 막대한 경제적, 혁신적 이익을 제공하지만, 동시에 의사 결정 과정에서의 차별, 정당한 절차, 투명성 및 이해 가능성을 보장하는데 위협을 제기하고 있다”(EU Civil Law Rules on Robotics(2017)).
- 사람이 갖고 있는 문화적인 경험과 문제점이 의도적이든 그렇지 않든 데이터를 통해, 프로그래밍을 통해 알고리즘에도 전이된다.

결과적으로, 알고리즘 의존도의 심화

- 알고리즘이 "인간에 대한 평가도구가 되고, 기본권을 제한할 수 있다"는 점이 확인되고, 의사결정지원을 넘어 사실상 의사결정시스템이 되면서 알고리즘에 대한 의존도가 높아지지만, 여전히 블랙박스 상태에 놓여있다.
- "알고리즘에 대한 인간의 의존도가 높아짐에 따라 그 공정성과 객관성에 대한 의문도 지속적으로 제기되고 있다"는 점은 알고리즘이 갖고 있는 본질적인 한계이자, 기술종속성에 대한 우려이기도 하다.

그래서, 알고리즘 투명성 논의

- 투명성 논의가 필요한 이유는 ”알고리즘이 단순한 계산에 그치지 않고 인간의 판단을 대신할 뿐 아니라 사회적, 경제적, 정치적 가치판단이 필요한 영역에까지도 확장하고 있기 때문”이다.
- 인공지능의 자율성이 커질 수록, 알고리즘에 대한 규제 입법이 이루어지고 있다. 다만, 논의의 방향은 기술 자체의 규제가 아닌 기술이 갖는 의미와 가치가 인간중심적으로 해석되고 적용될 수 있는 사회시스템의 설계에 있다.

알고리즘이 바꾸는 인간의 규범 : 기대 v. 우려 -> 규제?

- 알고리즘에 대한 대중의 기대는 “사람은 편향된 사고를 할 수 있지만, 기계는 객관적인 판단을 한다”는 것이다. 그렇지만, 인공지능에 기반한 판단의 기초가 되는 데이터 자체가 인간이 생산하고 입력하는 것이기 때문에 데이터의 한계가 그대로 결과로 나타나게 된다.
- 인공지능 기술의 혜택(惠澤)만큼이나 예상되는 부작용과 위험으로 인해 더 많은, 더 강한 규제의 필요성에 대한 압력도 높아질 수 있다.

영업비밀로서 알고리즘

- 영업비밀을 “공공연히 알려져 있지 아니하고 독립된 경제적 가치를 가지는 것으로서, 비밀로 관리된 생산방법, 판매방법, 그 밖에 영업활동에 유용한 기술상 또는 경영상의 정보”로 정의하고 있다.
- 영업비밀은 법적으로 구체화된 정보라기 보다는 사실상 비밀로 관리되는 정보이다. 특허와 같이 공개를 요건으로 하는 것이 아니므로, 그 실체를 확인할 수 없기 때문이다.
- 알고리즘도 이러한 요건하에 관리되는 경우라면, 영업비밀성 있기 때문에 영업비밀보호법의 보호대상이다.

왜, 알고리즘 공개는 필요한가?

- 블랙박스현상은 모델이 입력과 출력 사이의 복잡한 관계를 파악하기 위해 다양한 계층과 연산을 거치기 때문으로, 원인 중 하나는 딥 러닝 모델의 구조와 학습 알고리즘의 복잡성을 들 수 있다.
- 알고리즘 규제로서 공개가 추구하는 가치는 알고리즘에 의한 의사결정 과정의 투명성 확보, 신뢰성 확보, 시장에서의 알고리즘에 의한 교란행위의 차단 등 공정성 확보이다.

규제로서, 알고리즘 공개의 방향

- 알고리즘 규제는 현행 법제도를 기반으로 할 수 밖에 없으나, 알고리즘으로 인해 발생할 수 있는 피해 내지 예견되는 문제의 범위, 정도 및 미치는 영향에 따라 추가하거나 개선해 나가야 한다.
- 알고리즘 규제는 알고리즘에 따라 구현되는 가치와 이를 규제하고자 하는 보호법익의 비교형량에 의거하여 이루어져야 한다.
-

공개 대상 및 비밀유지 의무

- 알고리즘의 공개는 기본권의 침해나 개인의 법익의 침해 등 공익을 해하는 경우 등 예외적인 경우로 한정되어야 한다. 즉, 알고리즘이 사용되는 모든 경우에 해당하는 것이 아니라, 중대하게 공익이나 기본권을 훼손하는 경우 등 제한적으로 이루어져야 한다.
- 따라서, 알고리즘 공개 자체가 특정 다수 또는 불특정 다수 등 공중에게 이루어지는 것이 아닌 규제기관에 한정된다는 점을 명확히 해야 한다. 좀더 자세히는 규제기관의 해당 규제를 담당하는 특정인에게 공개하는 것으로 그 특정인은 비밀유지의무가 있는 자에 한정되어야 한다.

무엇을 공개해야 하나?

구분	대상	보호수준	대안
알고리즘	·소스코드 ·유사코드	소스코드에 한정하여 비밀유지의무 대상	설명가능AI, 사전영향평가
	·오픈소스	GPL 등에 따라 개작된 내용 공개	-
데이터	·학습데이터(구입, 자체구축 여부)	전체 학습데이터에 한정하여 비밀유지의무 대상	샘플 데이터 공개
	·오픈 데이터	오픈 데이터 조건에 따른 비밀유지	URL 등
처리 정보	·파라미터, 가중치, 정제 등 관련 정보	비밀유지의무 대상	업데이트 등 이력정보

EU AI법안에서의 공개

- EU AI법안은 데이터 등의 접근권한을 당국에 제공토록 하고 있다. 즉, 시장감시당국은 데이터 및 문서에 대한 액세스, 애플리케이션 프로그래밍 인터페이스('API') 또는 원격 액세스를 가능하게 하는 기타 적절한 기술적 수단 및 도구를 포함하여 제공자가 사용하는 교육, 검증 및 테스트 데이터 세트에 대한 완전한 액세스 권한을 부여받아야 한다.
- 또한, 고위험 AI 시스템이 제2장(고위험 인공지능 시스템의 요건)에 명시된 요건에 부합하는지 평가하는 데 필요한 경우, 합리적인 요청이 있는 경우, 시장감시당국은 AI시스템의 소스코드에 대한 접근권한을 부여받아야 한다. 물론, 국내 공공기관 또는 기구가 취득한 모든 정보 및 문서는 제70조에 명시된 기밀 유지 의무를 준수하여 취급되어야 한다(제64조).

공개에 따른 우려(문제)는 없는가?

- 기술중심론자들의 주장은 기술이 경제발전을 이끌어가기 때문에 함부로 기술 규제를 해서는 안된다는 것으로, 기업의 투자를 이끌어내기 위해서는 규제보다는 투자할 수 있는 환경을 마련해야한다고 주장한다.
- 알고리즘이 공개될 경우에 이를 악용하는 경우가 문제될 가능성이 있다. 예를 들면, 알고리즘이 공개된다면 알고리즘을 개선하려는 동기가 사라질 수 있다는 것이다. 특허의 경우, 기술은 공개하되 발명자에게 독점사용권을 주면서 동기부여를 제시한다. 그렇기 때문에 무조건적인 공개는 알고리즘 고도화 기회의 상실로 연결될 수 있다.
- 또한, 유사코드 형태의 기준이나 방법 등을 공개할 경우에는 이를 회피하는 경우도 예상된다.

우려를 불식시키는 것이 필요하다!

- 사업자들은 영업비밀인 알고리즘을 공개하는 것에 대해 부정적이나, 그 이유 중 하나는 알고리즘을 일반 공중에게 공개한다는 인식이 있다. 즉, 누구나 공개된 알고리즘에 접근하여 이용하거나 악용할 수 있다는 것이다.
- 그렇지만, 알고리즘 공개는 전문화되고, 접근이 제한된 공적 기구를 통하여 이루어지도록 한다는 점에서 영업비밀 유지의무가 있는 자로 한정될 것이라면 사업자의 우려는 불식시킬 수 있을 것이다.
- 한편, 공개의 기준을 제시하거나, 대체할 수 있는 방안을 찾는 것도 중요하다.

법에서 요구하는 '설명', 그리고 '의무'

- 알고리즘 설명 요구권(Algorithmic Explainability)은 인공지능 모델이나 머신러닝 알고리즘을 사용하여 어떤 결론이나 결과를 도출하는 과정에서 그 결과를 설명할 수 있는 권리를 말한다. 즉, 모델이 어떻게 작동하는지, 어떤 데이터를 기반으로 결론을 도출했는지, 어떤 요소가 결과에 영향을 미쳤는지 등을 설명할 수 있어야 한다는 것이다.
- 예를 들어, 의료 분야에서 인공지능 모델이 어떤 질병을 진단하거나 치료를 제안하는 경우, 해당 모델이 어떤 요소를 고려하여 그런 결론을 도출했는지를 설명할 수 있어야 신뢰성 있는 의사 결정이 이루어질 수 있다.
- 설명 요구권은 모델의 투명성과 신뢰성을 높이는 중요한 요소로 인식되고 있으며, 이를 충족시키기 위한 다양한 기술들이 연구되고 있다.

정보주체의 권리로서, 알고리즘 적용 거부권

- 알고리즘의 재산권 보장을 위해 정보주체에게 알고리즘의 구체적 내용의 공개를 요구할 권리를 부여하기는 어려우나, 알고리즘의 비공개 또는 제한적 공개가 정보주체의 기본권을 침해할 우려가 있을 경우, 정보주체의 방어권이 보장되어야 한다.
- 즉, 알고리즘의 제한적 공개에 대응해 정보주체에게 알고리즘 적용에 대한 거부권(拒否權)을 보장하는 것이다.

인공지능기본법에서의 신뢰성 확보방안

- 과학기술정보통신부장관은 인공지능등이 국민의 생활에 미치는 잠재적 위험을 최소화하고 안전한 인공지능의 이용을 위한 신뢰 기반을 조성하기 위하여 다음 각 호의 시책을 마련하여야 한다.
  1. 신뢰할 수 있는 인공지능 이용환경 조성
  2. 인공지능의 이용이 국민의 일상생활에 미치는 영향 등에 관한 전망과 예측 및 관련 법령·제도의 정비
  3. 인공지능의 신뢰성 확보를 위한 안전기술 및 인증기술의 개발 및 확산 지원
  4. 신뢰할 수 있는 인공지능사회 구현 및 인공지능윤리 실천을 위한 교육·홍보
  5. 인공지능사업자의 신뢰성 관련 자율적인 규약의 제정·시행 지원
  6. 인공지능사업자, 이용자 등으로 구성된 인공지능 관련 단체(이하 "단체등"이라 한다)의 인공지능의 신뢰성 증진을 위한 자율적인 협력, 윤리 제정 등 민간 활동의 지원 및 확산
  7. 그 밖에 인공지능의 신뢰성 확보를 위하여 과학기술정보통신부장관이 필요하다고 인정하는 사항
- 인공지능 및 인공지능기술을 개발·제작하는 자는 인공지능의 신뢰성 확보를 위하여 다음 각 호의 노력을 하여야 한다.
  1. 인공지능의 작동 과정과 결과 등이 논리적·객관적으로 설명이 가능한 기술개발
  2. 선량한 풍속 또는 국민 정서 등 사회적으로 수용할 수 있는 인공지능 기술개발
  3. 이용자의 기본적 권리의 보호

XAI, 설명요구권의 기술적 구현

- XAI는 인공지능 시스템의 의사 결정 과정을 설명하고 해석 가능한 방식으로 제시하는 기술이다. 인공지능 시스템의 의사 결정이 왜 그렇게 이루어졌는지 이해하는 것을 돕고, 투명성과 신뢰성 확보가 가능하다.
- 블랙박스 현상은 인공지능 시스템이 복잡한 모델을 사용하거나 대규모 데이터를 다루는 경우 특히 심해지는데, 이로 인해 인공지능 시스템이 내놓은 결과에 대한 신뢰성 문제가 발생하기도 한다.
- 따라서, 설명 가능한 인공지능은 블랙박스로 인한 문제를 해결하기 위해, 인공지능 시스템의 동작 원리를 사용자가 이해할 수 있는 방식으로 설명하고, 시스템의 결정 과정에서 어떤 요소들이 어떻게 작용했는지를 시각화한다.
- 이를 통해 사용자는 인공지능 시스템이 어떤 의사 결정을 내리는 데 사용된 근거를 이해할 수 있으며, 이를 바탕으로 신뢰성 있는 결정을 내릴 수 있다.

XAI에 대한 법적인 가치 평가

- XAI는 법에서 요구하는 설명요구권을 명시적으로 대체할 수 있다. 법적 의무를 이행함으로써, 알고리즘의 신뢰성을 확보할 수 있다.
- 알고리즘의 편향성을 해소할 수 있으며, 타당한 설명을 가능하게 한다.
- 알고리즘으로 발생할 수 있는 사고에 대한 원인, 책임 소재를 확인할 수 있으며, 손해배상의 책임범위를 설정할 수 있다.
- 잘못된 판단에 대한 원인 파악과 이에 따른 문제를 해결할 수 있는 방안을 강구할 수 있다.

EU AI법안의 기대와 우려

- AI법안은 EU의 다양한 논의를 통해 추진 중에 있으며, EU의 규제 주도권을 갖겠다는 의지도 내포되어 있다.
- 빠르게 움직이고 신속하게 용도 변경이 가능한 기술은 물론 규제하기 어렵다. 기술을 구축하는 회사조차도 상황이 어떻게 진행되는지 완전히 명확하지 않기 때문이다. 그러나 적절한 규제가 전혀 없이 운영을 계속하는 것은 분명히 우리 모두에게 더 나쁠 것이다(Ada Lovelace Institute).
- '미지의 기술'에 대한 환상보다는 대비가 필요하며, 알고리즘의 투명성을 요구하는 것은 그만큼 영향력이 크기 때문이다. 다만, 중소기업, 스타트업에 대해서는 예외로 한다.

‘알고리즘 공개 가이드라인’의 제안

- 알고리즘 공개에 대한 가이드라인을 제시할 경우, 사업자는 예측가능한 범위내에서 서비스를 개발하거나 제공할 수 있을 것이다. 이로써, 법적 안정성을 확보할 수 있을 것이다.
- 대략적인 가이드라인에 포함될 사항은 다음과 같다.
  - 알고리즘에 대한 기본이념
  - 문제를 해결하기 위한 방향성 제시
  - 데이터, 알고리즘 등의 공개 기준
  - 공개 방법
  - 비밀유지 의무 등의 구체화

감사합니다.