



## 웹포탈의 알고리즘은 그들의 욕망이다

네이버와 다음을 통해 제공되는 뉴스의 편향성이 화두다. MBC의 <스트레이트>에 따르면, 뉴스가 노출된 위치, 빈도, 기간 모두 수구언론사의 기사가 압도하고 있다. 이에 국내 웹포탈(Web portal) 업체인 다음과 네

이버는 인공지능(artificial intelligence)의 알고리즘(algorithm)이 한 일이라고 둘러댔다. 사람이 한 것이 아니라 컴퓨터가 한 일이니 공정하고 정확하다는 항변일 것이다. 박정희 시절 세금문제를 따지는 민원인에게 국제청 직원이 통명스럽게 “컴퓨터로 출력했다”며 훈계했다는 얘기가 떠오른다.

### 알고리즘이 바로 인간의 마음이다

그저 웃음이 나온다. 인공지능의 알고리즘이 어디 하늘에서 떨어지기도 했던 말인가? 이른바 기계학습(machine learning)이나 심층학습(deep learning or hierarchical learning)과 같은 아리송한 말로 사람들을 홀리고 있지만 진실은 간단하다. 신경망이든 양자 알고리즘이든 인간의 지능을 조금이라도 더 닮을 수 있을까, 어떻게 사람처럼 배울 수 있을까 하는 몹부름이다. 컴퓨터는 인간을 거울처럼 비치고 있다. 인공지능(알고리즘의 집합체)은 제작자의 마음과 의도를 고스란히 담고 있다. 아무리 처리속도가 빨라지고 저장능력이 커졌다 해도 사람이 만든 기계일 뿐이다. 인간의 마음과 의도를 반영하지 않는 기계는 아무리 우수해도 실패작이다. 부수어지고, 버려지고, 다른 기계에 끼워질 운명이다.

### 데이터가 운명을 좌우한다

둘째, AI는 철저하게 데이터에 의존한다(data-driven). 알고리즘의 논리구조에 따라 데이터를 분석하여 대상의 특성치(parameters)를 수정해 나간다. 이것이 학습이다. 따라서 규모가 크고 다양하고 믿을 만한 데이터는 좋은 결과를 낼 수 있다. 하지만 쓰레기를 입력하면 쓰레기가 나올 뿐이다. 아무리 슈퍼컴퓨터에 초대형 데이터(big data)를 넣어 돌린다 해도 마찬가지다. 전통적인 확률이나 회귀분석을  $P(\theta)$ 라고 표현한다면, Maximum Likelihood (ML)나 Bayesian 회귀 분석은 조건부 확률인  $P(\theta|y)$ 라고 말할 수 있다(King 1998: 14-18). 여기서  $y$ 는 데이터이고  $\theta$ (theta)는 특성치가 포함된 모델이다. AI는 주어진 데이터에서 특성치를 수정하면서 찾아가는 베이지안 방법을 취하고 있다. ML이든 베이지안이든 특성치의 운명은 전적으로 데이터에 달려있다.

### 네이버와 다음이 웃기고 자빠졌다

웹포탈의 인공지능이 바로 그들의 의지다. 알고리즘이 한 것이 아니라 바로 네이버와 다음이 한 것이다. 거울에 비친

자신에게 “니가 했지?”라고 손가락질하고 있다. 하지만 반드시 문제인 정권을 흔들고 말겠다는 웹포탈의 의도라기보다는 그들의 수익을 극대화하려는 욕망이다. 돈을 벌기 위해 클릭수와 노출시간에 매달릴 수밖에 없는 포탈의 셈법이다. 그들의 알고리즘은 이런 의도에 충실할 뿐이다. 이런 상황에서 사실과 상관없이 독자를 유혹하는 기사를 백미터 경주하듯 쏟아내야 하는 기자들의 압박이 있다. 기사를 쓰느라 취재할 시간이 없다는 황당한 푸념이다.

웹포탈의 인공지능이 뉴스내용을 이해하고 평가한다는 것은 불가능하다. AI의 학습은 진위를 판단하고 문맥을 파악하는 것이 아니라 패턴을 찾아 연결(추론)하는 것이다. 알고고가 바둑을 두는 것이 아닌 것처럼 AI는 기사를 읽는 것이 아니다. 컴퓨터는 단지 주어진 논리구조에서 계산을 pingpong히 빠르게 할 뿐이다. 사람들은 경이로운 계산능력을 지식능력으로 착각하고 있을 뿐이다. 둘째, 뉴스 자체가 새로운 것이기 때문에 과거 데이터에 의존하는 AI는 쓸모가 별로 없다. 주가를 예측한다는 AI가 허무맹랑한 까닭이다. 입력된 데이터가 없으면 나올 것이 없고, 적으면 정확성과 신뢰도가 떨어진다. 그래서 알고고가 이세들의 변칙에 속수무책일 수밖에 없다. 결국 웹포탈의 뉴스편집은 인공지능이 아닌 AI를 가장한 “인공지능”의 작품이다. 미끼상품을 골라 진열하는 일이다

정말 웹포탈이 인간의 역사, 정치, 사회, 문화, 종교 등을 이해하고 판단할 수 있는(더이상 데이터가 필요없이 정확한  $\theta$ 값을 알고 있는), 그러면서 엄청난 계산능력을 장착한 인간을 구현해 냈다면, 이 알고리즘은 인류 역사를 바꿀 수 있는 “제왕의 반지”다. 네이버와 다음이 불확실성(uncertainty)을 통제할 수 있다는 뜻이고, 세상을 지배한다는 의미다. 이자율과 주가를 예측할 수 있으니 온 세상의 뭉치돈은 다 그들의 차지가 될 것이다(주가예측 AI를 광고한다는 자체가 어불성설이다. 그런 AI는 뽕뽕 숨겨놓고 혼자만 사용해서 돈을 긁어모을 일이다). 대통령, 국회의원, 법관, 검사, 경찰, 종교인을 손바닥에 올려놓고 주무를 것이다. 네이버와 다음은 웹포탈 시장의 푼돈이 아니라 전 세계의 패권을 차지하기 위해 전쟁도 불사할 것이다. 하지만 이런 걱정은 불들어 매시라. 그들의 “수동 알고리즘”은 고작 문제인 정권을 흠집내고 낙인찍으려는 기사를 골라내고 있다. 한미정상회담을 해리스 부통령이 문대통령과 악수한 손을 담은 것으로 요약하는 수준이다. 앙증맞지 않은가.

### 네이버와 다음은 뉴스에서 손을 떼라

웹포탈과 언론사의 거래와 상술에 기자들은 노예가 되고, 시민들은 불량기사를 습관처럼 보면서 매출을 올려주고 있다. 클릭수와 노출시간에 쫓긴 기레기는 이제 모이를 쪼아대는 닭처럼 설사 없이 글쇠판을 두드리는 “계자鸚者”가 되었다. 수구세력이 지배하는 언론지형에서 AI의 편파성은 놀라운 것이 아니다. “Garbage in, garbage out”이라 했다. 시민들이 포탈이 골라주는 쓰레기를 생각없이 받아먹으면 거짓과 왜곡으로 꾸며진 메아리방(echo chamber)에 스스로를 가두게 된다. 허상에 중독되고 편파성이 강화된다. 사실 AI의 알고리즘을 공개하냐 마냐는 본질이 아니다. 포탈이 뉴스장사를 못하도록 해야 한다. 방송이든 신문이든 날조기사는 엄중히 처벌해서 퇴출시켜야 한다. 무엇보다도 깨어있는 시민 스스로가 진위를 따지고 진실을 밝히는 주체가 되어야 한다. 🙄

### 참고문헌

King, Gary. 1998. *Unifying Political Methodology*. Ann Arbor, MI: The University of Michigan Press.

**검**찰은 2019년 9월 6일 조국 교수의 법무부장관 인사청문회가 열리고 있는 시간에 배우자인 정경심 교수를 기소했다. 총장이름으로 나가는 표창장을 위조한 혐의지만 피의자를 조사하지도 않고 한밤중에 재판에 넘긴 것은 상식과 거리가 멀다. 검찰은 정교수에게 자본시장법과 금융실명법 위반 등 14개 혐의를 씌워 기어코 구치소로 보냈다. 언론보도에 비친 조국 내외는 반역죄를 저지르고도 반성하지 않는 파렴치한이다. 하지만 “차고도 넘친다”는 검찰의 증거는 1심 재판에서 대부분 허무하게 무너졌다. 다른 것은 몰라도 총장 표창장 위조에 관한 검찰의 설명과 법원의 판단은 한마디로 어이가 없다. 사실에 근거한 증거 제시와 과학적인 추론이라 할 수 없다.

### IP주소가 뭐길래?

검찰은 동양대에서 압수한 PC에 저장된 IP (Internet Protocol) 주소를 근거로 정교수가 2013년 6월 방배동 자택에서 표창장을 위조했다고 주장했다. IP주소는 네트워크에 연결된 장치(예컨대, 컴퓨터, 프린터, 스마트폰, 라우터)를 인식하기 위해 장치에 부여되는 번호다. IPv4는 8비트(0~255) 네 개를 붙여서 만든 번호체계로 최대 43억개 주소를 정의할 수 있다. IP주소는 특정 번호로 고정될 수도 있고(포트에 번호가 지정되어 있고), DHCP (Dynamic Host Configuration Protocol)가 그때그때 다른 번호를 할당할 수도 있다. 서버를 운영하려면 고정 IP주소가 필요하고, 대부분(정교수의 PC 포함)은 효율성이 높은 변동 IP주소를 사용한다.

ICANN에서 관리하는 공인 IP주소는 전체 인터넷에서 고유한 번호로서 Public 혹은 Global 주소라고 하고, 개별 네트워크(흔히 LAN)에서만 고유한 IP주소는 Private 주소라고 한다. 정교수가 라우터(Router)를 통해 PC를 연결했다고 하는데, 이는 라우터의 DHCP에서 사설 IP주소를 할당받았다는 뜻이다. 검찰은 192.168.123.xxx라고 했다. 정교수가 LGU+ 제품을 사용한 듯하다.

### IP주소로는 위치를 알 수 없다

인터넷에 연결되는 장치가 많아짐에 따라 IPv4로는 감당하지 못하고, 16비트 여덟개를 붙인 IPv6는 호환성 문제로 널리 사용되지 못하고 있다. 이 상황에서 IP주소를 전환해주는 NAT (Network Address Translation) 기술이 대안이 되었다. 라우터의 NAT는 PC에서 외부로 신호(패킷)를 보낼 때 사설 IP주소를 공인 IP주소(라우터의 IP주소)로 변환하고, 외부에서 PC로 정보를 받을 때는 공인 주소를 사설 주소로 바꾸어준다. 따라서 똑같은 사설 IP주소가 동시에 다른 네트워크에서 사용될 수 있다. 한 네트워크 안

## 정경심 PC의 IP주소와 알리바이

에서만 고유한 번호라면 상관없다. 결론적으로 변동 사설 IP주소로는 PC의 위치를 특정할 수 없다. 검찰이든 변호인이든 마찬가지다. 내 PC의 사설 IP주소는 192.168.50.xxx이다. 여긴 방배동도 동양대도 아니다. 지리적 위치와 무관하게 제조사(Asus)가 정한 주소체계일 뿐이다. 이 라우터를 사용하는 한 세계 어디에서 접속하든 같은 IP주소(범위)를 받는다. 게다가 IP주소를 변조하여 서버를 속이는 방법도 있다.

### MAC 주소는 좀 다른가?

검찰은 윈도우를 재설치하기 전과 후의 MAC (Media Access Control) 주소가 같으니까 PC가 방배동에 있었다. 황당하다. MAC주소는 제조사에서 네트워크장비(Network Interface Controller)에 물리적으로 적어놓은 고유번호다. MAC주소가 같다는 것은 그 PC의 네트워크카드를 빼내지 않았다는 뜻이지, 방배동인지 아닌지와는 무관하다.

그러면 어떤 조건에서 IP주소나 MAC주소로 장치의 위치를 알 수 있을까? 첫째, 잘 관리된 고정 IP주소를 사용해야 한다. IP주소를 지정하는 부서에 가면 정확한 위치를 확인할 수 있다. 둘째, 서버가 라우터의 공인 IP주소, 할당된 변동 IP주소, MAC주소를 같이 저장해놓았다면, 특정한 시간대 어느 장치가 어느 변동 IP주소로 작업했는지 알 수 있다. 하지만 이런 네트워크 환경을 가진 경우는 흔하지 않다.<sup>1</sup> 인터넷서비스회사(ISP)나 구멍가게 수준에서 이런 정보를 관리할 가능성(수익성)은 거의 없다.

컴퓨터를 좀 한다는 사람이라면 아는 사실을 검찰이 몰랐을 까닭이 없다. 이런 사실을 확인하지 않고 검찰의 억측을 받아들인 1심 재판부의 저의가 의심스럽다. 정말 방배동 주소가 적혀있는 줄로만 믿었을까?

### 최성해의 거짓말이 증거다

인사청문회가 열리기 하루 전(참으로 절묘한 시기다) 검찰에 출석한 당시 동양대 학교 총장 최성해씨는 조국 교수의 딸에게 표창장을 발급한 적이 없다고 말했다. 일정한 일련번호를 부여받아야 하고 발급기록이 보관되기 때문에 총장 자신도 모르게 표창장이 발부될 수 없다고 했다.

<sup>1</sup> 인디애나대학교에서 컴퓨터를 네트워크에 연결하면 서버는 컴퓨터의 MAC주소를 읽어서 사용자 데이터베이스에 등록되어 있는지를 확인한다. 목록에 없으면 사용규정에 동의하고 10분 이상 기다렸다가 다시 접속하라고 한다. 서버가 누가, 언제, 어느 장치(MAC주소)를, 어느 변동 IP주소로, 어디서(대강의 위치), 얼마동안 사용했는지를 기록하고 있다는 얘기다.

검찰청을 나오면서 해맑게 교육자의 양심을 선택했다는 최씨를 보면서 나는 혀를 찼다. 정말 기억력이 좋아서 수년 전 발급된 표창장까지 다 꿰고 있던 말인가? 총장이 일련번호가 어떻게 생겼는지, 발급절차가 어찌되는지, 어떻게 기록이 보관되는지 시시콜콜 다 알고 있었던 말인가? 학과나 연구소에서 요청하는 표창장이 얼마나 많을 것인가? 일련번호와 형식은 물론이고 기록을 관리하는 일도 쉬운 일이 아니다. 실제로 일련번호 형식이 서로 다른 동양대 총장 표창장이 여러 개가 발견되었다.

뉴스에 나온 최씨를 보면서 나는 2007년 학력위조로 세상을 떠들썩하게 했던 신정아씨를 떠올렸다. 그녀는 공학에서 출국하면서 예일대 박사학위를 증명하겠다고며 기업을 도했다. 나는 아연쩍었다. “아니 어떻게 본인 스스로가 박사학위를 증명한다 말인가? 지도교수와 심사위원이라면 모를까... 정말 박사과정 공부를 한 적이 없는 사람이군...” 최씨도 학위 5개 중 3개가 가짜로 드러났다. “교육학박사 최성해”가 찍힌 표창장 자체가 허위라는 것 아닌가.

### 가짜박사와 법쟁이들의 비양심

가짜박사의 인터뷰를 본 소감은 (1) 총장 노릇을 제대로 못한 사람이다(동양대의 학사가 엉망이다), (2) 당시 정황으로 보면 최씨는 정교수의 딸에게 표창장을 발급해 주었는지 아닌지를 잘 알고 있다, (3) 발급해 주었는데도 아니라고 말했을 가능성이 크다. 평소 머느리감이라며 애지중지하던 사람에게 표창장이 아니라 명예박사인들 아깝겠는가. 지금은 관계가 틀어졌지만서도.

정말 최씨가 표창장에 대해 몰랐다면 일련번호나 장부를 운운하지 않았을 것이다. 검찰에 가기 전에 살펴보고 계산을 했다는 뜻이다. “교육자의 양심”은 1999년 옷로비 사건 때 “정경에 손을 얹고 맹세할 수 있다”는 말을 떠올리게 한다. 양심있는 사람은 무겁게 진실을 말할 뿐 속보이는 소리를 하지 않는다. 최소주의와 거리가 멀다.

나는 표창장이 위조되었는지 아닌지 모른다. 다만 검찰과 최씨의 주장이 사리에 맞지 않다는 것이다. 이런 식으로 기소하고, 사실과 과학은 외면하고, 증인과 언론을 동원하여 망치를 두드린다면 판검사 편에 서지 않고서는 살아남을 수가 없다. 동양대 표창장이 뭐길래 이 난리인가? 과연 한 집안을 풍비박산 낼 일인가? 법쟁이와 가짜들의 양심에 모진 털이 무성하고, 백성들의 이성과 상식은 아파서 울고 있다.

