Law Super AI의 필요성

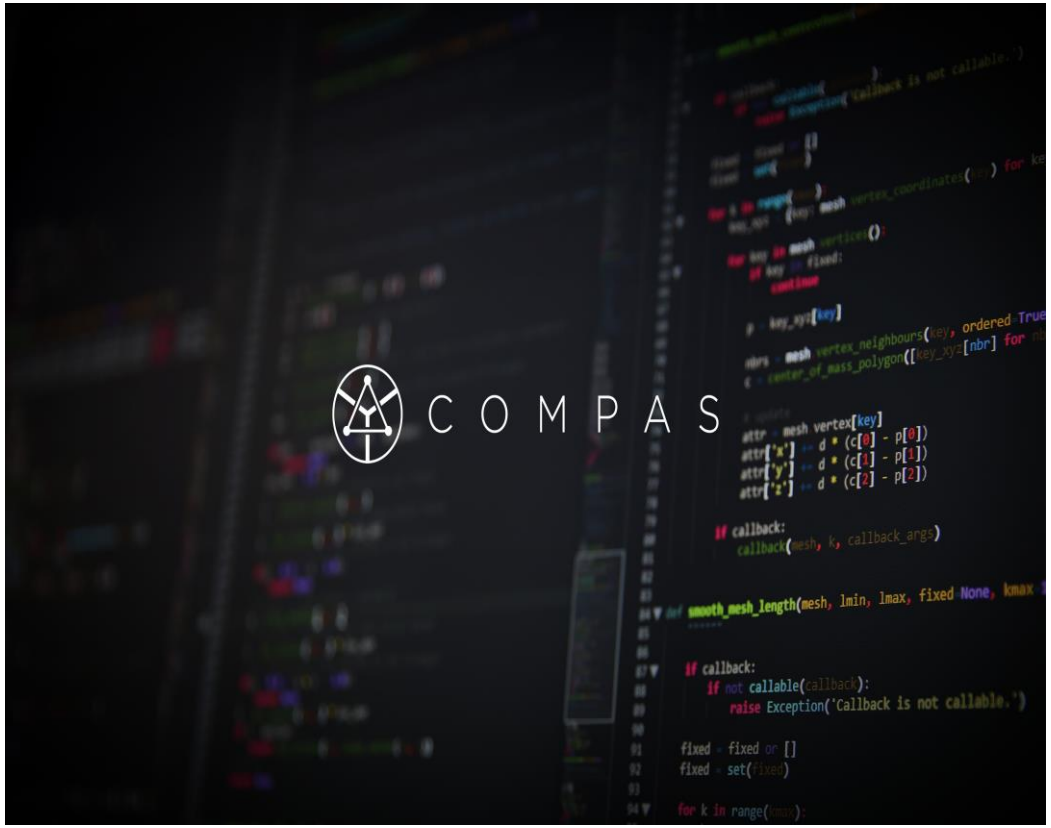
목차.

서론 : 인공지능은 차별을 할 것인가?

1. 논문 리뷰 : 2016년

* 컴파스 알고리즘 : 2016년 버전이 백인 vs 아프리카계 미국인들에 대한 차별이 분명히 있다고 언급을 했고, 실제로 위험점수가 높은데 실제로 2년 안에 재범이 일어나지 않을 확률은 백인이 23.5%, 아프리카계 미국인 44.9% 로 약 2배 가량 차이가 난다며, 위험점수가 낮은데 실제로 2년 안에 재범이 일어날 확률은 백인이 47.7%, 아프리카계 미국인이 28% 를 기록했다. 그러나 다른 주장에 따르면 59% – 63%, 71% – 69% 라고 언급이 될 정도로, 이 부분에선 전혀 다른 차이를 보여줬다라고 한다.

* 미국 대법원에 판결 사례에서 보듯이 기본 인종비율이 다르다. 그래서인지 몰라도, 빅데이터에 의존해 판단한 사례라고 볼 수 있다. 그렇게 된다면 알고리즘 같은 경우는 당연히 흑인 리스크가 있을 것이며, 그에 따른 리스크 점수로 사람을 평가하게 될 것이다. 리스크 점수가 흑인에게 더 높을 가능성이 높으며, 그에 따른 알고리즘이 나온 상황이고, 그렇게 학습을 한다면 알고리즘은 똑같은 범죄를 저지르더라도 흑인에게 더 높은 점수를 부여해, 형벌을 가할 것이다.



(사진 : 컴파스 회사)

2. 그리고 시간이 지났다.

* 컴파스 vs 인간
재범 확률 예측의 정확도 : 인간 (흑인 68.2%, 백인 67.6%), 컴파스 (흑인 64.9%, 백인 65.7%)
흑인에 대한 위양성의 비율도 : 인간 (흑인 37.1%, 백인 27.2%), 컴파스 (흑인 40.4%, 백인 25.4%)
백인에 대한 위음성의 비율도 : 인간 (흑인 29.2%, 백인 40.3%), 컴파스 (흑인 30.9%, 백인 47.9%)
=〉 아직은 모든 경우에 인간의 판단이 조금은 우위에 있는 상태이다.

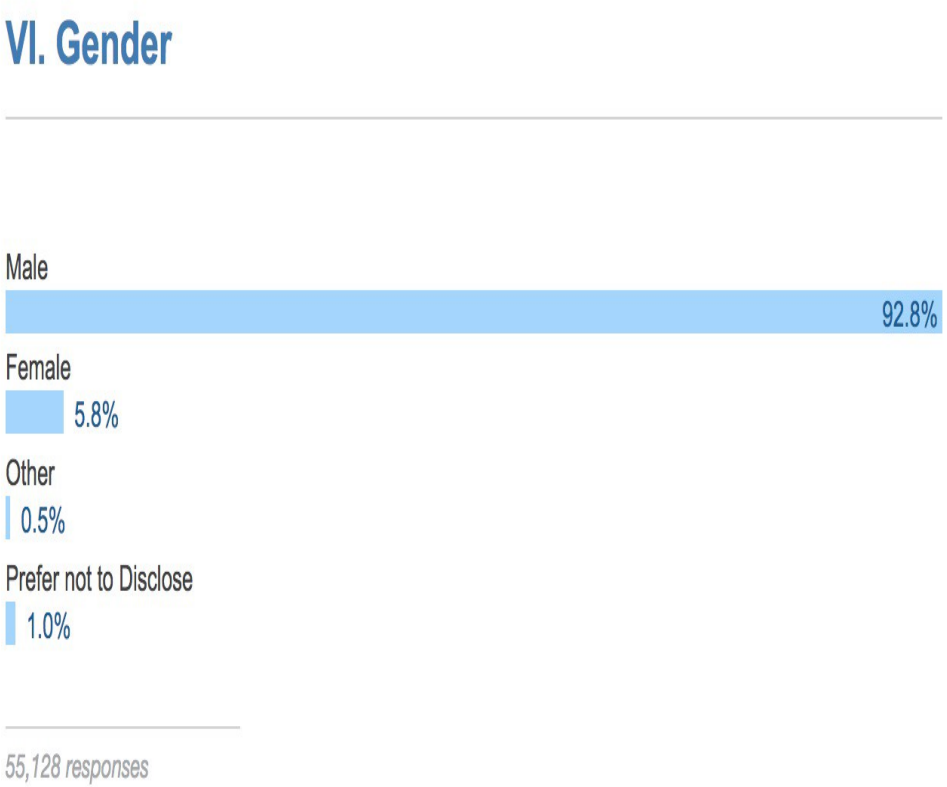
* 그런데 '아직은' 이다. 시간이 지나면 알파벳이든, 텐센트든 움직이지 않을까. 인간보다 더 뛰어난 알고리즘이 나오지 않을까 생각이 된다. 아직은 이기고 있다고 하지만 바둑도 2014년에 비해 5년 뒤에 발전 되었던 상황인데, 이것도 2021년이면 사람이 알고리즘에게 거의 원사이드로 질 가능성이 있지 않을까 생각된다.

(사진 : 여러 약 인공지능들)

3. 여러 알고리즘에 대한 문제 (1) – 성비율, 인종비율

* 개발자의 성비율이 2016년 기준으로 남자 92.8%, 여자 5.8%, 기타 1.4% 이다. 이게 무슨 의미인지 아는가? 남자 위주로 돌아간다는 것이 현실이다. 알고리즘은 개발자의 손을 거친 다는 것을 잊어서는 안된다. 이런 상황이라면 미투운동이 빈번하게 벌어질 수도 있다. 인공지능이 페미니즘을 잘 모르고, 공평한 물건이 아니라고. 안그런가?

* 유럽/백인 비율이 2016년 기준으로 74.2%, 남아시아 11.5%, 라틴 6.7%, 동아시아 5.1%, 중동 4.1%, 흑인 2.8%, 그외 0.8% (원주민등)로 나와있다. 백인 남성 비율이 굉장히 높은 편이다. 그렇게 된다면 어쩌면 그들도 모르게 그런 인식으로 가있는 상황에서 학습을 한다면? 당연히 백인 위주, 남성 위주로 가있지 않을까? 그렇게 된다면 알고리즘은? 어쩌면 백인들 위주로 편성이 될 가능성이 있다. 실제 범죄 데이터에서도 똑같은 범죄를 저질러도 혜택 (예를 들어서 형량을 내린다던가) 을 저지른 사례가 있다. 그렇게 된다면 알고리즘은 인식을 못한다. 똑같은 범죄에 대해서 인간이 형량을 제멋대로 하는데, 과연 그런걸 학습한 알고리즘이 범죄 지수로 매긴다면 당연히 지금까지의 인공지능은 차별을 하는 것이다.



(사진 : 2016년 남녀 개발자 비율)

4. 여러 알고리즘에 대한 문제 – (2) 트롤링

* 극단주의 (ex) 네오나치, 인종차별주의, 극단적 남, 여성 우월주의)가 딥러닝을 한다면? 당연히 인공지능에 대한 차별이 있을 것이다. 당연하다. 혹은 그런 사상을 가진 사람이 실제론 아니어도 장난을 치는 행위, 장난 이상의 깽판, 즉 트롤링을 한다면? 상황은 제멋대로 변할 가능성이 높을 것이다.

* 또한 알고리즘은 상황에 따라 변할 수도 있다고 생각한다. 질서 선이 될 수도 있고, 중립 선, 혼돈 선이 될 수도 있다. 질서 중립, 중립, 혼돈 중립, 질서 악, 중립 악, 혼돈 악이 될 수도 있다. 즉 군자가 될 수도 있고, 자선가가 될 수도 있고, 다크 히어로로, 판관, 나무수염 (반지의 제왕), 잭스패로우, 카이사르, 무한이기추의자, 조커보다 더욱 미친 알고리즘이 될 수도 있다. 문제는 이 9개 알고리즘이 매초마다 변할 가능성이 농후하다.

* 이렇게 된다면 당연히 인공지능은 사람을 차별하는 것이 된다. 트롤링을 하는 것을 막을 순 없기 때문이다. 안 그런가? 알고리즘이 사람을 통해서 만든 것이기 때문에 어쩔 수가 없다.

(사진 : 반지의 제왕의 나무수염. 중립)

본론 : 인공지능이 차별을 안하려면?

1. 약인공지능의 업그레이드

* 지금까지 모든 알고리즘은 약인공지능이었음. 약인공지능이란 특정한 분야의 일을 인간의 지시에 따라 업그레이드 하기 위한 연구가 필요함. 그게 어떤 것이든 상관없이 리스크 점수를 더욱 수정을 해야 할 필요가 있다고 생각한다. 적어도 모든 사람에게 조금이라도 더 공평한 사람이라고 하면 각나라에 모든 법을 넣고, 상황에 따른 시뮬레이션을 넣어서 알파고 제로 이상의 무언가를 만들어야 할 필요가 있지 않을까?

* 조금 더 나아가보면, Psycho-pass 의 시빌라 시스템처럼 (Psycho-pass는 애니메이션이다) 범죄계수를 만들어서 개인, 단체간의 점수를 해 일정 점수가 넘어가면 자동적으로 벌을 내리는 시스템 (이 시스템은 범죄계수가 300이 넘어가는 순간 즉결사형을 시킨다!)을 만들되, 점수가 떨어지지 않는 시스템을 만든다.

* 시빌라 시스템은, 모든 공소시효를 없애는 시스템을 만드는 것이다. 물론 이 가능성은 2110년대를 배경으로 만들었기 때문에 아직까지 근미래의 이야기이기도 하고, 아직 만화속에서만 가능한 이야기이긴하다.

사진 : Psycho-pass 의 시빌라 시스템

2. 대안 – 초지능 개발?

* 각 나라, 각 지역의 법을 1초안에 흡수를 할 가능성이 높기 때문에 아무래도 그런 페널티를 자기 스스로 알아서 점수를 매길 가능성이 높지 않을까. 그렇게 된다면 적어도 불공정을 막을 수 있기 때문이다. 또한 그런 칩을 달게 된다면 자연스럽게 범죄 지수 (1명 죽이면 몇점, 부정 입학 몇점, 2천억 달러 회계비리면 몇점.)를 세팅을 해서 최대 몇 점이면 그에 따른 처벌하는 시스템이 나오지 않을까?

* 그런 지수를 만들어버린다면 두가지 장점이 있다. 하나는 법이 최대한 공정해진다는 점이다. 또한 사람들은 승소를 하려고 김앤장이나 그런 곳에다가 돈을 마구 뿌릴 필요가 없을 것이다. 단점은 변호사, 검사, 판사들은 일자리가 없어진 다는 점이다.

결론 : Law Super AI를 개발해야 한다.

1. Law Super AI를 개발하는 이유?

* 약인공지능보다 더욱 발달이 된 것이 필요하고, 그리고 더 뛰어난 초지능의 개발이 필요한 상황이다. 아니. 더 중요하게 보자면 초지능 이상을 개발하는 것이 목표다. 왜냐하면 인간 보다 더 뛰어난 초지능, 그 초지능을 완벽한 중립으로 할 수 있는 시스템을 만들어야지 차별을 하지 않을 것이다.

* 각 나라별로 법이 공평하지 않으며, 사람마다 공평하지 않기 때문에, 그런 법에 따른 Scoring System 을 만들어야 한다. 즉 Law Super AI가 나와야 한다. Law Super AI에 따른 일정 점수대로 법을 처리해야 한다면 지금보다 더욱 공정한 일들이 벌어지지 않을까. 적어도 형량을 조작할 순 없기 때문이다. 왜냐하면 Law Super AI를 만드면, 거기에 따른 룰을 따르기 때문이다.

* 점수는 만든 시점부터 적용해야 한다. 그리고 타임머신이 개발이 된다면 과거에 따른 점수도 매겨야 할 것이다.

* 그렇게 된다면 Law Super AI의 개발은 성공적일 것이다.

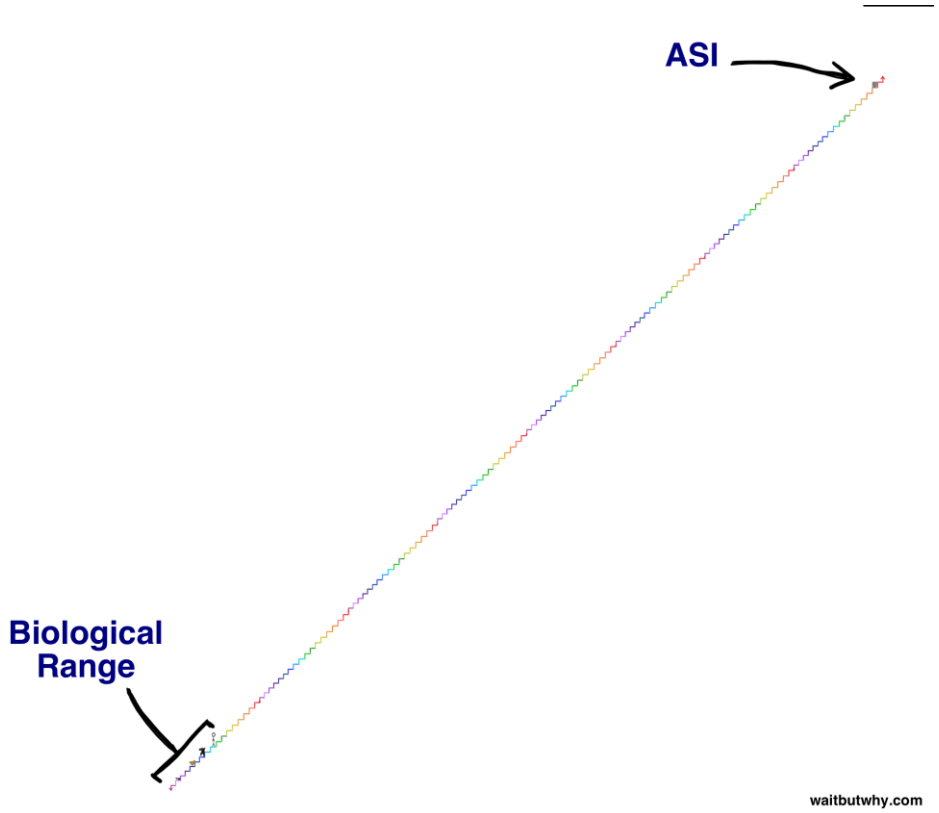
(사진 : 인공지능의 발전 방향)

2. Law Super AI 프로젝트

* Law Super AI 는 초지능이니까 사람들에게 모두 연결을 해버림과 동시에 범죄 지수를 측정하는 것이고, 일단 0부터 측정을 해서 일정 점수가 넘어가면 처벌을 하는 것이다. 모든 것들을 사람이 판단이 하는 것이 아니라 Law Super AI가 판단하는 것이다.

* 각 모든 사람, 모든 생물체에게 적용이 되며, 태어날 때부터 죽을 때까지 적용이 되는 것이다. 모든 것은 Law Super AI 가 판단하도록 내버려 두는 것이다. Law Super AI 가 시스템이 파괴되지 않도록 여러 개를 만드는 것이다.

* 물론 지금 이 시점에서의 가능성은 먼 미래의 이야기일지도 모른다. 하지만 적어도 만들 가치가 있다면, 더 나아가 인간보다 훨씬 더 판단을 잘한다면? 적어도 서둘러서 만들어야 하지 않을까? 그것이 얼마이든 간에 만들어야 한다. 6G, 7G 시대에 대비해야 하니까 말이다.



(사진 : 초지능과 생물체의 차이)