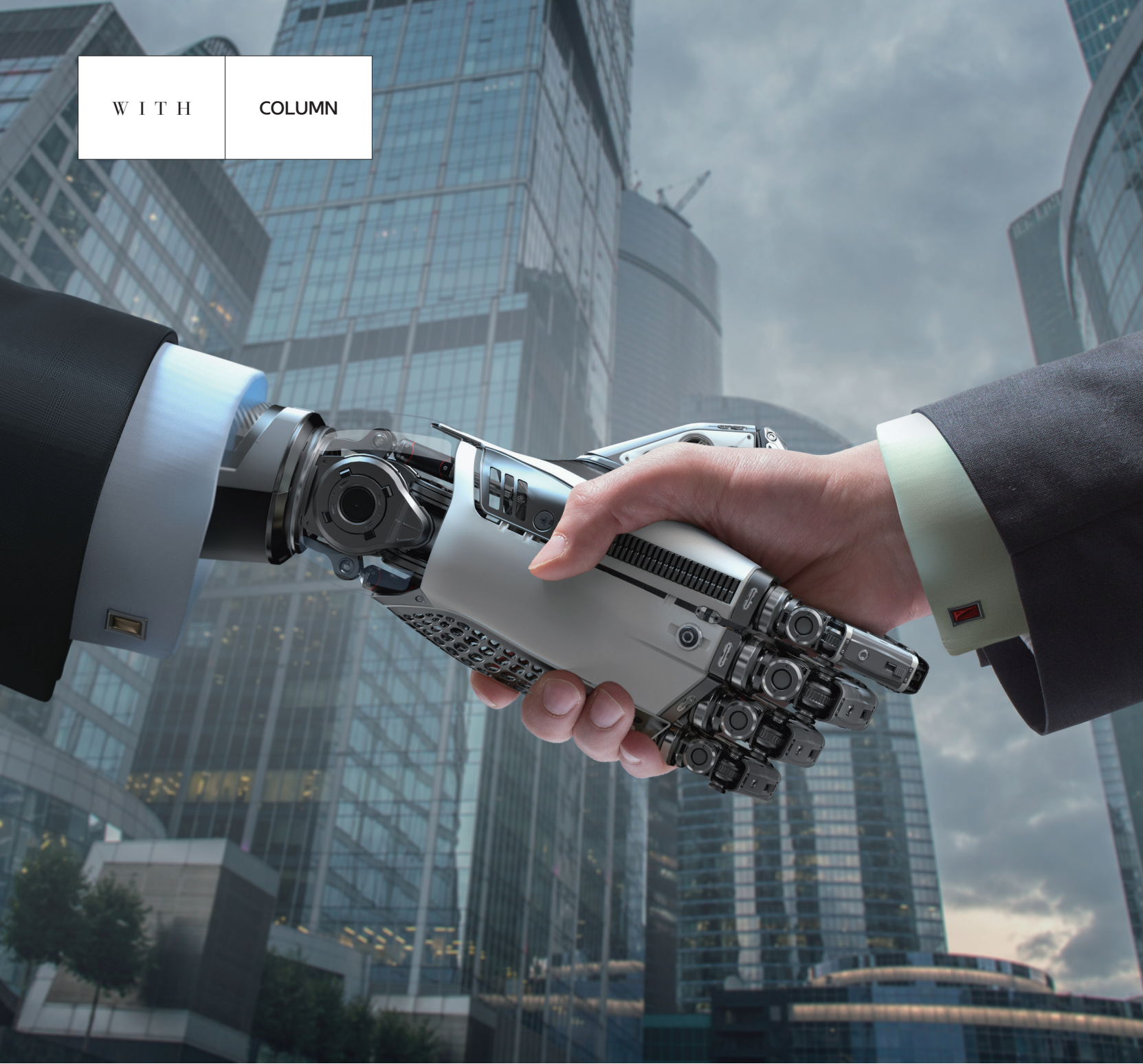


WITH

COLUMN



## AI와의 협업과 AI 공유를 통한 번영

인간과 인공지능(AI)은 둘 다 불완전하다. 따라서 불완전한 인간은 역시 불완전한 AI와 협력해 자신을 발전시킬 수밖에 없다. 그런데 AI 개발을 위해서는 데이터가 필요하고, 개인과 조직의 존재 의미를 위해서는 데이터와 프라이버시의 보호가 필요하다. 데이터를 보호하면서도 이를 활용해 다양한 AI 모델을 개발, 공유하는 체계를 촉진시키는 한편 이러한 사용자 중심의 AI를 공유하고, AI를 통한 온 좋은 발견을 해나가는 방식으로 인류는 자신의 건강, 지식, 행복, 문화를 발전시켜 나갈 것이다.



**이경전**

[경희대학교 교수 & 하렉스인포텍  
사용자중심인공지능연구소 소장]

▶ 경희대 경영대 & 빅데이터융합학과 교수  
경희대 빅데이터연구소 소장  
AI & BM Lab 지도교수

## AI의 정의

AI는 말 그대로 인간이 만드는 지능이다. 인간을 닮은 것도, 인간을 넘어서는 것도 아니다. 그럼 지능은? 제임스 앨버스는 1991년 발표한 'Outline for a Theory of Intelligence'에서 성공적인 행동을 산출하는 것을 지능으로 정의한다. 생각하는 것이 아니라 행동을 산출하는 것이다. 그리고 성공이란 목표를 달성하는 것이다. 따라서 우리 인간은 AI 시스템에 그 목표를 잘 설정해줘야 한다. 그럼 목표를 달성한다는 것은 무엇일까? 목표를 달성하기 위해서는 합리적으로 적절히 행동해야 한다. 수학적으로는 목표 함수를 최적화하는 것이다.

## 인간도, 컴퓨터도 불완전하다

그런데 AI는 불완전(Incomplete, Imperfect)하다. 인간 역시 완벽하게 합리적일 수 없다. 인간의 합리성은 제한적이다. 인간이 지니는 계산 능력, 의사결정 능력의 한계가 존재하기 때문이다. 허버트 사이먼은 인간에 대한 솔직한 관찰과 설명을 한 후 제한된 합리성이라는 개념을 내놓으며 노벨경제학상과 튜링 어워드를 수상했다.

NP-Hard, NP-Complete 이론(NP·Non-Polynomial) 발전에 큰 기여를 한 스테픈 쿡과 리처드 카프는 컴퓨터가 아무리 발전해도 세상의 많은 문제는 우리가 원하는 시간에 발견하기 어렵다는 일종의 불가능성에 대한 발견으로 컴퓨터 과학 분야의 노벨상인 튜링 어워드를 받았다. 쿼트 컴퓨팅이 이를 극복하기 위한 하나의 방편으로 연구되고 있지만 이 역시 NP-Hard, NP-Complete 이론 자체를 극복하는 것은 아니고 단지 답을 구하는 시간을 단축하는 방법론일 뿐이다.

경제학자 프리드리히 하이에크는 엘리트끼리 모여 중앙집중적으로 의사결정하는 방식으로는 좋은 사회를 운용할 수 없다는, 이른바 '중앙집중적 계산불가능성'을 일찍이 간파했고, 그의 경제사상과 철학으로 노벨경제학상을 수상했다. 결국 서로 교환, 협상, 흥정하는 시장 메커니즘이 필요하다는 의미로 해석될 수 있다. 하이에크는 지금 유행하고 있는 딥러닝, 뇌과학의 발전을 예견한 듯한 '감각적 질서 : 이론 심리학의 이론적 기초에 대한 탐구'라는 독특한 저서를 남기기도 했다.

쿠르트 괴델은 어떠한 공리체계도 증명할 수 없는 참인 명제가 항

상 존재하며, 따라서 스스로 모순성이 없음에 대한 증명은 불가능하다는 불완전성 정리를 증명했다. 물리학 분야 노벨상 수상자 베르너 하이젠베르크는 어떤 입자의 위치와 운동량을 모두 정확하게 알 수는 없다는 불확정성의 원리를 발견해 인류를 더욱 겸손하게 만들었다. 세상 만물의 어떤 상태를 정확하게 파악할 수는 없다는 의미로 해석된다. 케네스 애로는 우리가 바람직하다고 여기는 의사결정의 기준을 모두 만족시키는 의사결정 메커니즘은 만들 수 없다는 것을 증명한 불가능성 정리를 제시해 노벨경제학상을 받았다.

요즘 유행하는 블록체인 역시 블록체인 트릴레마(Trilemma)라는 것이 있다. 이더리움을 발명한 비탈릭 부테린이 제시한 용어로 확장성(Scalability), 보안성(Security), 탈중앙화(Decentralization)를 모두 만족시키는 블록체인을 만들기는 어렵다는 것이다. 일견 멋져 보이는 블록체인이 아직도 사회에 큰 가치를 제공하고 있지 못하는 이유다. 이 모든 발견은 인간을 겸손하게 하고, AI의 한계에 대해 생각하게 만든다.

## AI는 영원히 실수할 것이다

AI는 실수를 한다. 최적의 해를 구하려고 노력했지만, 실제 최적의 해가 아닌 경우 AI는 실수를 한 것이 된다. AI는 영원히 실수를 할 것이다. 인간이 늘 실수를 하는 것과 비슷하고, 기계가 늘 고장나고 오류를 일으키는 것과 비슷하다. 우리 인류는 AI가 실수를 한다는 점을 인식하고 그 점을 오히려 활용해야 한다.

## 불완전한 인류가 실수하는 불완전한 AI를 사용하는 방법

필자는 이 글에서 불완전한 인류가 실수하는 불완전한 AI를 사용하는 방법을 네 가지로 나누어 소개하고자 한다.

첫 번째는 세렌디피티(Serendipity)적 활용이고, 두 번째는 여러 AI 모델을 결합하는 앙상블, 세 번째는 디지털 미(Digital Me), 마지막은 AI 공유다.

필자가 연구소장으로 활동하고 있는 하렉스인포텍 사용자중심 인공지능연구소가 개발한 음식 추천 AI 시스템이 어느 날 고객에게 로제치킨 세트, 제육토스트, 낙지크림 스파게티를 추천했다. 이러한 상품은 세상에 존재하지 않는 제품이었다. AI가 실수를 한 것이다. 그런데 달리 보면 AI는 실수를 한 것이 아니라 신상품

W I T H

AI가 존재하지 않는 상품을 추천 : 실수(?) vs 신상품 아이디어(?)

19855269



오리지널 치킨



땡초 김밥

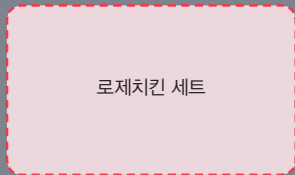


만두



오리지널 치킨

- 사례
- 제육토스트
  - 낙지크림 스파게티
  - ...
  - ...



로제치킨 세트

유사 상품 탐색



로제떡볶이

'오늘 먹을 음식 추천'

<그림 1> 자연어 처리 기반의 추천 시스템을 활용한 상품 기획 아이디어 도출 방법론  
출처 : Brainstorming AI(이경천, 정백, 황보유정, 조영재, 2022 KMS Spring Conference)

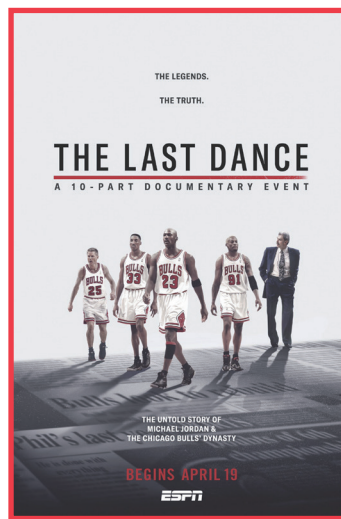
아이디어를 인간에게 준 것이다. 흥미로운 것은 로제치킨 세트는 치킨 프랜차이즈 회사 BBQ가 그 사이에 출시했다.

세렌디피티 : 운 좋은 발견

이렇듯 위험도가 낮은 분야에 AI를 사용하면, AI의 실수가 세렌디피티가 돼 운 좋은 발견을 할 수도 있다. 필자가 제일 좋아하는 단어 중 하나가 세렌디피티다. 세렌디피티는 신성과 인간성 사이, 우연과 필연 사이, 행운과 노력 사이, 마술과 공학 사이에 있다. 루이 파스퇴르 박사는 'Prepared Mind', 즉 준비된 자에게만 세렌디피티가 온다고 했다(Chance favors only the Prepared Mind!). 여기서 준비란 필연을 만드는 인간의 노력인 공학을 의미한다. 그러한 공학은 우연, 마술, 행운, 신성처럼 느껴지는 세렌디피티를 창출할 수 있다.

Many Model Approach

실수가 많은 AI를 잘 활용하는 또 하나의 방법은 여러 AI 모델을 검토하고, 그것을 융합하는 것이다. 필자는 금융 분야의 AI를 개발할 때 여러 모델을 개발한 후 그 모델들이 내놓은 결과를 잘 조



화시키는 방식을 활용했다. 이렇게 하나의 문제에 대해 여러 모델을 검토하는 것을 'Many Model Approach'라 할 수 있다.

AI 시스템 개발과 운영은 농구단, 야구단을 경영하는 것과 비슷하다. 우수한 농구선수 한 명을 키우는 것도 중요하지만 우수한 농구팀을 이끄는 감독, 그런 팀을 만드는

구단주의 마인드로 AI 시스템을 만들어야 한다. 전성기의 시카고 불스를 묘사한 다큐멘터리 'Last Dance'의 교훈이기도 하다. 선발 투수(모델)만이 아닌 불펜 투수(모델)를 운용하고, 후보 및 신인 투수(모델)가 계속 공급될 수 있도록 지속적인 모델 수립 및 검토를 진행하는 팀을 운영해야 한다. 원활한 후보 모델을 도입하기 위해 AI 시스템을 모듈화하고 여러 기법을 조화시키는 방법론을 활용해야 한다.

SCOTT E. PAGE

# THE MODEL THINKER

Why “Many-Model Thinkers” Make Better Decisions

by Scott E. Page

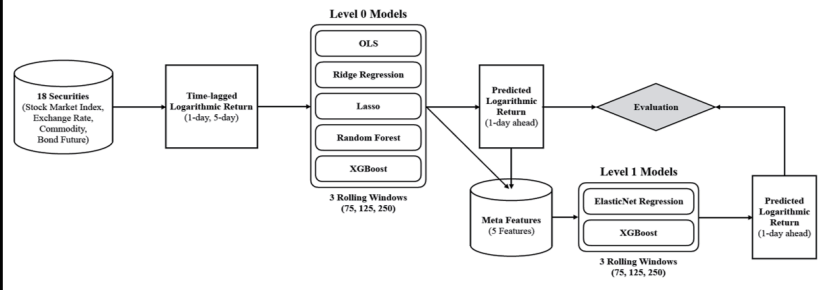
Harvard Business Review

November 19, 2018

## Bankruptcy Prediction Modeling Using **Multiple** Neural Network Models

Kyung-shik Shin, Kyoung Jun Lee · Published in KES 2004 · Computer Science

The primary goal of this paper is to get over the limitations of single neural network models through model integration so as to increase the accuracy of bankruptcy prediction. We take the closeness of the output value to either 0 or 1 as the model's confidence in its prediction as to whether or not a company is going to bankrupt. In case where multiple models yield conflicting prediction results, our integrated model takes the output value of the highest confidence as the final output. The output of the confidence-based integration approach significantly increases the prediction performance. The results of composite prediction suggest that the proposed approach will offer improved performance in business classification problems by integrating case-specific knowledge with the confidence information and general knowledge with the multi-layer perceptron's generalization capability. [LESS](#)



〈그림 2〉 금융시장 예측을 위한 앙상블 접근  
출처 : 한재윤, 이경전, 2017 KIIS

### 사례와 지식의 크로스체킹

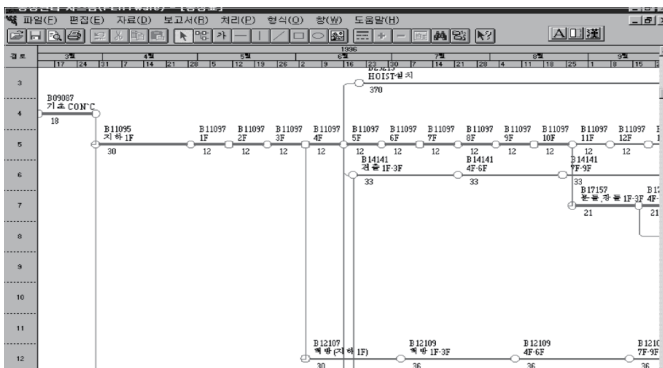
아파트나 교량과 같은 건축물 설치 계획을 하는 AI를 만든 경험이 있다. 새로운 계획을 세울 때 우리는 과거의 사례와 현재의 지식을 결합한다. 같은 방법으로 AI 시스템을 개발했다. 잘했던 과거 사례에서 현재를 다듬어 나가고, 현재의 첨단 지식을 수집하는 등 사례와 원칙을 크로스체킹하며 발전시켰다. 사례에 위배되는 지식을 개선하고, 지식에 위배되는 사례를 개선했다. 그렇게 함으로써 AI가 인간의 지능을 보완하고 인간이 AI를 보완하는 협동 시스템을 만들었다.

### 상호 보완과 점진적 개선

이렇게 불완전한 인간과 불완전한 AI를 상호 보완하게 하는 체계를 구축하는 것이 중요하다. 어떤 하나의 완벽한 인간지능, 모든 것을 다 잘하는 AI를 개발하는 것이 아니라 사회에 존재하는 불완전한 인간과 불완전한 AI 시스템이 서로의 지능을 공유하고 협력하는 것이 더 중요하다.

이것은 세계관과도 연결되는데, 혁명을 꿈꾸기보다는 결국 점진적으로 개선해 나가는 것밖에 방법이 없다는 의미로 해석될 수 있다. 점진적으로 개선하기 위해 인간은 목표를 설정하고, 데이터와 AI의 도움을 받아 가장 비슷한 사례와 비교해 그 사례가 잘못된 방향으로 문제를 풀어갈 수 있다.

필자는 어렸을 때 탁구를 좋아했는데, 필자보다 탁구를 더 잘치는 친구를 발견하면 매일 학교 정문에서 그 친구를 기다렸다. 그리고 그 친구에게 탁구비 내기 게임을 제안했다. 첫째 날과 둘째 날은 지고 탁구비를 냈지만, 일주일 정도 지나니 필자가 그 친구를 이길 수 있게 됐다.



〈그림 3〉 아파트 건설을 위한 사례 기반 및 제약 기반 프로젝트 계획

### 인간과 AI가 서로를 투영

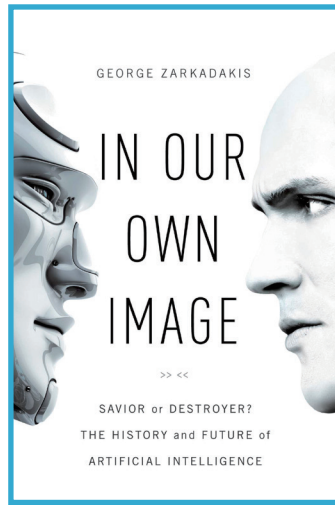
필자는 이렇게 불완전한 존재인 인간과 불완전한 AI가 서로를 서로에게 투영해가면서 스스로를 점진적으로 개선해 나갈 수 있다고 생각하며 연구하고 있다. 비유한다면, 자신의 자화상을 AI가 계속 그려주고 보여줘서 인간의 발전을 도와주는 것이라고 할 수 있다.

그리스의 AI 박사 조지 자카다키스의 'In Our

Own Image'라는 명저는 인간은 자신이 발전시킨, 자신이 창조한 문명에 자기 자신을 투영해왔다고 설명한다. 성경의 창세기는 인간을 흙에 비유했다. 왜 그랬을까? 창세기가 쓰일 당시는 인류의 문명이 흙의 문명이었기 때문이라는 것이다. 인간이 태엽시계를 발명했을 때, 인간은 우리 스스로를 태엽시계로 생각했다는 것이다. 이러한 관점에서 보면, 요즘 AI가 결국 사람이 될 것이고, 인간은 디지털하다는 생각도 결국 인간이 자신이 만든 디지털과 AI 문명에 자기 자신을 투영한 오류라고 볼 수 있는 것이다.

인류는 끊임없이 자신을 물과 거울에 투영해왔다. 동화 '백설공주'의 왕비도 그랬고, 나르키소스는 연못에 비친 자신을 사랑하게 되면서 결국 연못에 빠져 죽었다.

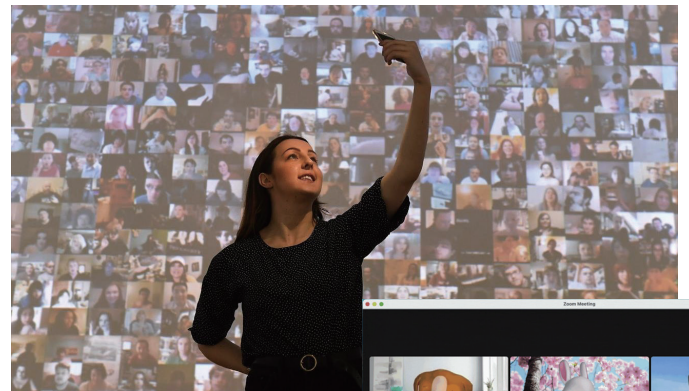
예술가 역시 끊임없이 자화상을 그려왔다. 김재이 작가가 이쾌대의 자화상을 패러디한 자화상이 있다. 이쾌대의 얼굴이 있던 자리에 김재이 작가의 얼굴이 있고, 그 모습을 바라보는 작가의 뒷모습과 결합한 멋진 작품이다. 이쾌대에 대한 존경과 시간의 거스름을 표현하고 있다. 오스카 와일드의 소설 '도리안 그레이의 초상'에서는 젊고 아름다운 도리안 그레이의 초상이 그가 타락함에 따



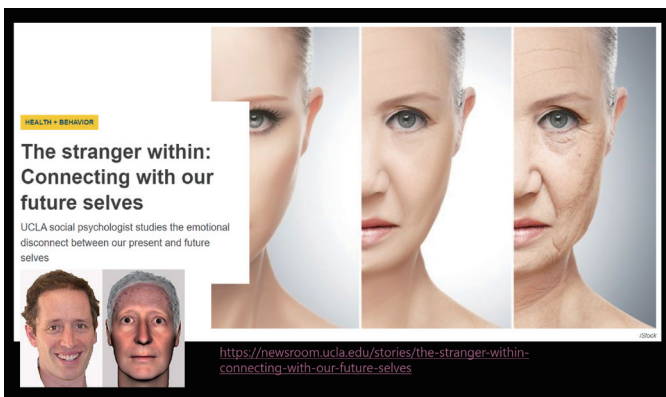
라 같이 추해진다. 우리 인간은 자신이 타락하고 추해지고 있는 것을 알면서도 자신이 순수했고 아름다웠던 시절을 기억하기 위해 자신만의 초상을 유지하고 싶어 한다. 아름다운 초상이 자신의 추함을 가려주길 원하는 것이다. 그러나 와일드는 그 초상마저 추해지는 상상력을 동원해 타락해가는 도리안 그레이를 더욱 크나큰 고통에 빠뜨린다.

수십 편의 자화상을 남긴 프리다 칼로는 평생 고통을 주었던 그의 육체를 보여주는 동시에 그의 자유로운 정신과 꿈을 표현한다. 프랜시스 베이컨의 자화상은 자기모멸을 보여주는 대표적 예술작품이다. 라디오헤드의 'Creep' 속 가사와도 일맥상통한다. 우리는 자기모멸의 자화상을 그린 베이컨을 보며 작은 위로를 받는다. 누구나 자기모멸을 하는구나. Creep을 부르며 스스로 위로받는 것과 같다. I'm a creep, I'm a weirdo.

2017년 영국 런던의 사치 갤러리에서 셀피에 관한 특별 전시가 열렸다. 사람들은 셀카로 자기 자신을 바라보기도 하고, 줌과 같은 일종의 메타버스 서비스를 사용할 때는 아바타를 활용해 자신을 더 자유롭게 꾸미기도 한다.



미국 로스앤젤레스 캘리포니아대(UCLA) 헬 허시필드 교수와 마이크로소프트(MS)의 공동 연구에 따르면 일종의 AI 프로그램을 활용해 사람들에게 미래의 늙은 모습을 보여줬을 때, 사람들은 미래에 대비하면서 갑자기 운동을 하고 보험에 가입했다고 한다.



## 디지털 나(Digital Me)

필자는 우리 인류가 AI를 활용해 자신의 자화상을 그리고 이를 통해 자신을 관리하고 발전시키는 것이 현명하다고 생각한다. 생물학적 인간이 디지털 인간을 만들어 거기에 자신을 투영하고 자신을 추동하는 것인데, 이름을 붙인다면 디지털 나(Digital Me)라고 할 수 있다.

자신과 모습이나 건강상태, 살아온 궤적, 취향 등이 상당히 비슷한 사람을 우연히 만나면 어떻게 할 것인가? 어떤 사람은 “꼬옥 껴안아줄 것”이라고 한다. “그동안 많이 힘들었지?” 하면서.

필자는 인생이 허무할 것 같다는 생각을 했다. 자신만의 유일성, 유니크함을 개발했다고 생각했는데 어느 날 모든 것이 나와 비슷한 사람이 또 있다면 허무할 것 같았다. 그런데 지금은 생각이 바뀌었다. 어차피 자신과 비슷한 사람은 어딘가에 있다. 밀란 쿤데라가 이미 그러한 이야기를 했다. 그의 책 ‘지혜’는 “인간은 태어날 때 수백만 개의 육체 중 한 개를 받게 되는데, 큰 호텔의 수백만 객실 중 하나를 할당받는 것과 같다. 육체는 우연적이고 비개성적인 것, 즉 잠시 빌려 쓰는 일회용품에 불과하다”고 말했다. 프랑수아 리카르는 쿤데라의 희곡 ‘자크와 그의 주인’에 대한 평론에서 유일한 것이란 덧이고, 불행은 강박적 차이 추구에서 오며, 독창성은 허상이고, 청소년기의 천진한 소산일 뿐이며, 일종의 자만이라고 역설한다. 진정한 자유와 지혜는 반복에 대한 자각에서 온다는 것이다. 여기서 지혜를 빌린다면, 만약 자신과 비슷한 상태의 타인을 만난다면 자신보다 더 나은 것을 발견하고 이를 자신에게 도입하는 삶을 사는 것이 현명할 수 있겠다는 생각을 하는 것이다.

그런데 디지털, 네트워크, 데이터, 그리고 AI는 세상 어딘가에 있는, 또는 있었던, 또는 앞으로 있을 자신과 유사한 타인을 발견하는 것을 돕고 이에 기반해 자신의 건강, 지식, 행복, 문화적 감성 등을 발전시키는 것을 도울 수 있다.

## 데이터 공유가 아니라 AI 공유

그런데 이러한 AI를 개발하려면 개인의 데이터가 필요하고 이를 활용해 기계가 학습을 해야 한다. 그런데 개인의 데이터는 개인의 것이다. 빅테크 플랫폼이 그 개인의 데이터를 다 가져간다면 인간은 사생활이 사라지고 고유의 인간성이 다 없어지게 된다. 결국 빅테크 플랫폼이 개인의 데이터를 가져가는 것은 최소화해야 한다.

기업이나 정부 등의 조직도 조직의 존재 의미를 위해 데이터를 보호하는 것은 당연하다. 최근 개인과 기업의 데이터를 서로 공유하지 않고도 AI를 학습시키는 방법론이 개발됐다. 이러한 연합 학습 기술을 이용해 사회 주체 간 AI를 공유하는 것이 가능해졌다. 개인의 데이터를 수집하거나 개인의 데이터를 교환하는 것이 아닌, AI를 교환하고 공유하는 것이다.

요즘 책임성 있는 AI에 대한 논의가 많은데, 필자는 이를 더욱 적극적으로 수행하는 사용자 중심 AI에 관심을 두고 연구해왔다. 사용자 중심으로 AI를 연구하다 보니 사용자의 데이터를 다 가져와서 AI를 만드는 것이 아니라, 반대로 AI를 만들어 사용자에게 주고 사용자가 자신의 데이터를 가지고 AI를 학습시킨 후 그 AI를 사용자 간에 공유하는 AI 공유에 기반한 비즈니스 모델이 앞으로 크게 발전하게 될 것이라는 신념을 갖게 됐다. 데이터는 각 주체가 소유·유지하게 하고, AI는 상호 간에 공유하게 함으로써 성과는 높아지고 비용은 낮아지게 된다. 그렇게 하면 데이터를 노출하거나 공유하지 않고도 충분한 성과를 내는 AI를 개발할 수 있다.

에덤 스미스는 교환이라는 것은 인간이 동물과 구별되는 특성이란 설명하면서, 사람이 집에서 혼자 핀을 만들면 하루에 20개도 못 만들지만 열 명이 분업해 만들면 1인당 하루 4800개도 만들 수 있다고 설파했다. AI도 혼자 만들면 별 볼 일 없는 것을 만들게 되지만, 서로 공유하면서 발전시켜 나간다면 우리는 진정으로 사용자 중심, 인간 중심의 AI를 만들 수 있을 것이다.

더글러스 노스는 재산권을 잘 보호하면서도 그 활용에 있어 거래 비용을 줄이는 사회가 번영한다는 것을 실증적으로 밝혀 노벨경제학상을 받았다. 데이터도, AI도 결국은 자산이고 중요한 재산권이다. 따라서 이 재산권인 데이터와 AI를 잘 보호하면서도 이들의 활용에 있어 거래비용을 줄이는 신뢰 주체, 즉 AI 공유체계가 발전한다면 우리 사회와 개인도 더욱 발전하게 될 것이다.