

2019 Vol.01

차세대리포트

인간처럼 보고 생각하고 예측하라

머신러닝

KAST 한국과학기술원
The Korean Academy of Science and Technology

차세대리포트 2019 Vol.01

머신 러닝

인간처럼 보고 생각하고 예측하라

펴낸곳

한국과학기술한림원
031)726-7900

펴낸이

한 민 구

발행일

2019년 10월

홈페이지

www.kast.or.kr

기획·편집

배승철 한림원 정책연구팀 팀장
이동원 한림원 정책연구팀 주임
조은영 한림원 정책연구팀 주임
박주이 한림원 정책연구팀 주임

디자인·인쇄

경성문화사
02)786-2999

이 보고서는 복권기금 및 과학기술진흥기금의 지원을 통해 제작되었으며,
모든 저작권은 한국과학기술한림원에 있습니다.

발간사

지난해 한국과학기술한림원은 젊은 과학자들의 생각과 아이디어를 담은 정책제안서인 ‘차세대리포트’를 새롭게 발간하였습니다. ‘젊은 과학자를 위한 R&D 정책’, ‘과학자가 되고 싶은 나라를 만드는 방법’ 등의 주제를 통해 우수한 젊은 과학기술인들의 그룹인 ‘한국차세대과학기술한림원(Young Academy of Science and Technology, Y-KAST)’ 회원들의 목소리를 담았으며, 이를 통해 정책관계자들에게 새로운 정책의 실마리를 제공하고 과학기술계에서 소통과 공감을 이끌어내는데 조금이나마 일조하고자 했습니다.

특히 ‘영아카데미, 한국과학의 더 나은 미래를 위한 엔진’을 주제로 발간된 차세대리포트 2018-04호에서는 Y-KAST로 모인 젊은 과학기술인들은 어떤 목적을 향해 무슨 일들을 해나가야 하는지에 대한 설문조사와 소규모 인터뷰를 진행하였습니다. 매우 흥미롭게도 이들은 성별, 전공분야 등과 상관없이 두 가지에 대해 일치된 목소리를 내었습니다. Y-KAST는 정부와 젊은 연구자들 사이의 통로로 기능해야 하며 차세대회원들은 후속세대 양성에 대한 깊은 책임감으로 미래를 위한 나침반이 되어야 한다고 입을 모아 이야기한 것입니다.

올해 발간되는 차세대리포트는 이들이 제시한 두 개의 목표를 향해 나아갈 수 있는 발판이 되었으면 합니다. 본 보고서를 통해 미래 핵심기술(머신러닝, 양자기술, 수소경제)의 연구개발 이슈와 함께 해당 분야에서 우리나라가 확실한 우위를 선점하기 위해서는 어떤 정책적 지원이 필요한지 살펴보고, 이공계 인재들이 진로를 선택하는 데 있어 도움이 될 만한 생생한 정보를 전달하고자 합니다.

그 시작으로 차세대리포트 제1호에서는 우리의 일상 전체를 최적화하며 영향력을 키워가고 있는 머신러닝에 대해 다뤄보고자 합니다. 지난 15년 사이에 10⁴배 늘어난 기사가 쏟아지고 있는 머신러닝이 무엇인지에서부터 해당 분야를 선도하기 위해 필요한 과학기술정책까지 현장에서 직접 연구하고 있는 젊은 과학자들의 생각과 아이디어를 담았습니다. 이들이 전하는 목소리가 일반 대중을 비롯하여 국가의 미래에 대해 치열하게 고민하고 있는 우리의 젊은 세대들과 정책입안자들에게 작은 도움이 될 수 있기를 기대합니다.

함께해주신 분들



가천대학교
약학과 교수

김미현

신규 약물의 설계·합성, 새로운 합성법 개발, 구조-활성 관계(SAR)로 연결되지 않는 활성 절벽(activity cliff) 규명 등 산업 약학 기초 연구를 데이터 마이닝 기법으로 수행하고 있다. 머신러닝이 연구과정뿐만 아니라 의사결정 구조, 협업의 방식 등에 얼마나 큰 영향을 주고 있는지에 대해 상세히 이야기하고 우리나라의 신약개발 연구가 더욱 효과적으로 이뤄지기 위한 아이디어를 구체적으로 제시했다.

아주대학교
산업공학과 교수

이성주

기업 경영에서의 내비게이션이라 불리며 미래를 결정짓는데 중요한 역할을 수행하는 비즈니스 인텔리전스 분야에서 탁월한 연구성과를 선보이고 있다. 한국기술혁신학회 신인연구자상 수상에 힘입어 기술평가와 예측, R&D 협력 등에서 학술연구를 주도하고 있다. 머신러닝의 발전으로 해당 분야의 판도가 어떻게 바뀌고 있는지 이해하기 쉽게 설명해주었으며 이런 변화의 시기에 인간은 어떻게 대응해야 하는지 등에 대해 시사점을 제공했다.



서울대학교
수리과학부 교수

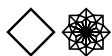
현동훈

대수기하학적 기법을 컴퓨터 비전, 특히 3차원 영상 재구성에 응용해 팜 테크(Farm Tech), 무인 판매 등 다양한 신성장 산업에 활용할 수 있도록 앞장서고 있다. 차세대리포트를 통해 컴퓨터비전과 머신러닝 기술의 오늘날을 이야기하고 앞으로 나아가야 할 전망을 제시하는 데 적극적인 의견을 내주었다. 더불어 수학자로서 기초학문과 타 응용분야와의 융합의 중요성을 강조하며 머신러닝 인력양성 방안에 대한 아이디어를 제시했다.



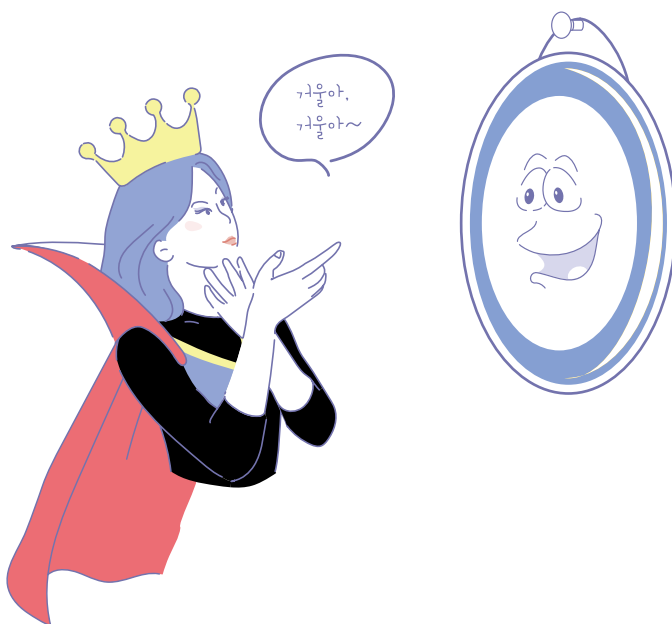


들어가기	06
의제 1	컴퓨터 비전·의약·비즈니스 분야의 머신러닝 개발과 적용 현황
1. 컴퓨터 비전과 머신러닝, 컴퓨터에게 보는 법을 가르치다	10
2. 머신러닝, 신약개발의 신(新)이 되다	15
3. 미래를 향하는 나침반, 비즈니스 인텔리전스	19
의제 2	머신러닝과 함께 나아가야 할 길
1. 우수 인재 확보에 머신러닝의 내일이 달려있다	23
2. 차세대 과학자, 정책을 이야기하다	26



들어가기

스스로 보고 생각하고 예측하는 ‘머신러닝’



“백설공주의 미모를 시기한 왕비가 마법 거울 앞에 섰다. “거울아, 거울아, 세상에서 누가 가장 예쁘니?” 그러자 거울이 말했다. “‘예쁘다’는 말의 뜻은 무엇인가요? 조건(특징값)을 구체적으로 수치화해 입력하면 순위를 정해드리겠습니다.”

지난 2018년, IT 엔지니어를 위한 추천 도서로 널리 각광 받은 <백설공주 거울과 인공지능 이야기 : 볼츠만 머신러닝에서 딥러닝까지>의 한 대목이다. 무엇이든 답해준다는 마법 거울의 비결이 인공지능(Artificial Intelligence, 이하 ‘AI’)이라는 은유는 유쾌하면서 자못 의미심장하다.

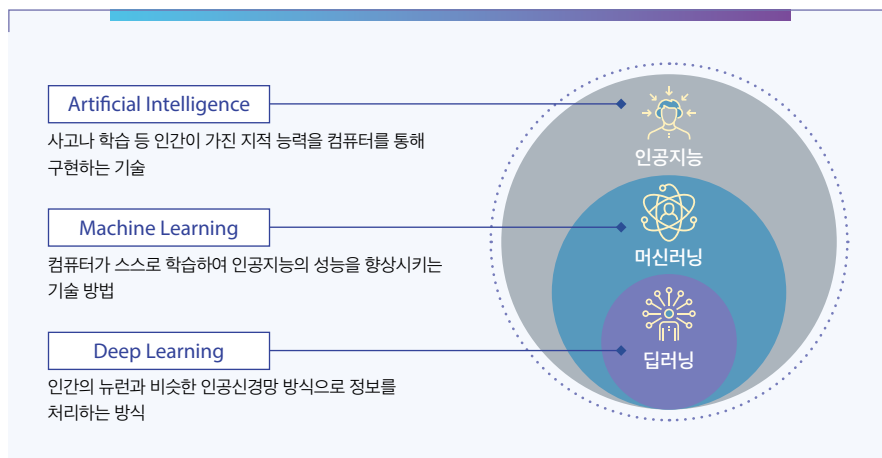
우리가 잘 알고 있는 아마존의 경우 매출 중 35%는 알고리즘이 추천하는 상품에서 창출된다. 또한 버락 오바마는 2012년 재선캠프 때 머신러닝을 활용하여 유권자들의 인구통계를 분석했다. 18~24세 남성들이 화요일에 즐겨보는 애니메이션에 광고를 해 투표를 독려하는 정교한 선거운동을 펼친 것이다. 그리고 페이스북의 Moments 서비스는 촬영한 인물을 기준으로 사진을 자동으로 분류해주며, 사이버 보안업체 캐스피다는 머신러닝을 통해 공격자의 행동패턴을 분석하여 이상징후를 탐지한다. 이렇게 우리는 매일매일 인공지능, 그 중에서도 머신러닝(Machine Learning, 기계학습)을 통한 결과물을 접하고 있다.

그렇다면 머신러닝이란 도대체 무엇이기에 이렇게 우리 삶 전반에 영향력을 나타내고 있는 것일까? 우리가 그 개념을 제대로 알지 못한 채 엄청난 변화의 흐름에 몸을 맡겨버린다면 마법 거울을 제대로 쓸 줄 모르면서 자신이 원하는 답을 듣길 바라는 왕비와 다를 바 없을 것이다. 따라서 그간 수박 겉핥기로 들어온 머신러닝의 본질을 알기 위해 노력해 볼 필요가 있다.

머신러닝의 선구자 중의 한 명인 아서 사무엘(Arthur Samuel)은 머신러닝을 “컴퓨터에게 명시적으로 프로그래밍 하지 않고 학습을 할 수 있는 능력을 주는 연구 분야”로 정의했으며 그것은 마치 아이와 같다고 비유했다. 아이가 아무것도 모르는 채 세상에 태어나서 그들이 경험한 것을 바탕으로 세상을 이해하게 되는 것과 마찬가지로 머신러닝은 인간이 입력한 프로그램이 아니라 경험한 데이터에 따라서 학습하고 스스로 반응을 결정하기 때문이다.

결국 머신러닝은 인공지능이 ‘인간의 지능을 기계로 구현’하고자 하는 것을 가능케 해주는 학습방법 중 하나로서 인간은 많은 데이터와 함께 최소한의 모델을 제공해야 한다. 이 모델에는 결정트리학습법, 귀납논리계획법 등 여러 가지가 있는데 이 중 사람의 뇌를 본 뜬(인공신경망) 구조를 가진 머신러닝 방식이 바로 딥러닝(Deep Learning)이다. 다만 ‘인공지능’은 굉장히 광범위한 의미를 가지고 있으며, 딥러닝만으로 연구 활용범위를 한정짓는 것은 한계가 있다고 판단되어 본 보고서에서는 머신러닝을 중점적으로 다루고자 한다.

◆ [그림 1] AI · 머신러닝 · 딥러닝의 관계





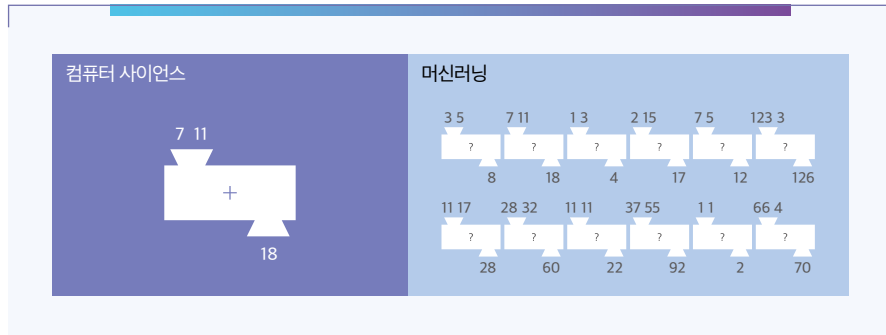
들어가기

기계학습(머신러닝)은 로켓의 엔진과 같다. 로켓이 날아가려면

“엔진에 넣을 연료가 필요한데 이것이 바로 데이터이고, 데이터는 IoT를 이용한 센서에서 얻어진다.”



[그림 2] 기존 컴퓨터 사이언스와 머신러닝의 차이점



많은 전문가들은 머신러닝이 여태까지 보여왔던 변화의 양상보다 더욱 획기적인 변화를 우리의 삶 속으로 불러일으킬 것이라고 전망하고 있다. 에릭 슈미트(Eric Schmidt) 전 구글 회장은 “5년 내에 모든 기업이 머신러닝을 사용할 것이다”라고 말했으며 시장 조사기관들은 특히 2018~2020년 사이 연평균 50% 이상의 고속 성장을 기록할 것으로 예측하고 있다. 세계적인 컨설팅 기업 맥킨지는 오는 2025년 머신러닝을 포함한 AI 기술 발달로 인한 사회경제적 파급 효과가 연간 최대 6조 7,000억 달러(한화 약 8,000조 원)에 이를 것이라고 내다보고 있으니 앞으로 우리의 삶이 얼마나 더 바뀔지는 쉽게 가늠하기 어려울 정도다. 이에 한국과학기술한림원은 과학기술계 연구현장의 최전선에 있는 Y-KAST 회원(이하 차세대회원) 3명을 통해 머신러닝 적용 분야의 대표주자인 컴퓨터 비전·신약 개발·비즈니스 인텔리전스에 대한 구체적이고 실질적인 이야기를 들어보았다.

AI산업은 이제 막 혁명의 입구에 들어섰으며

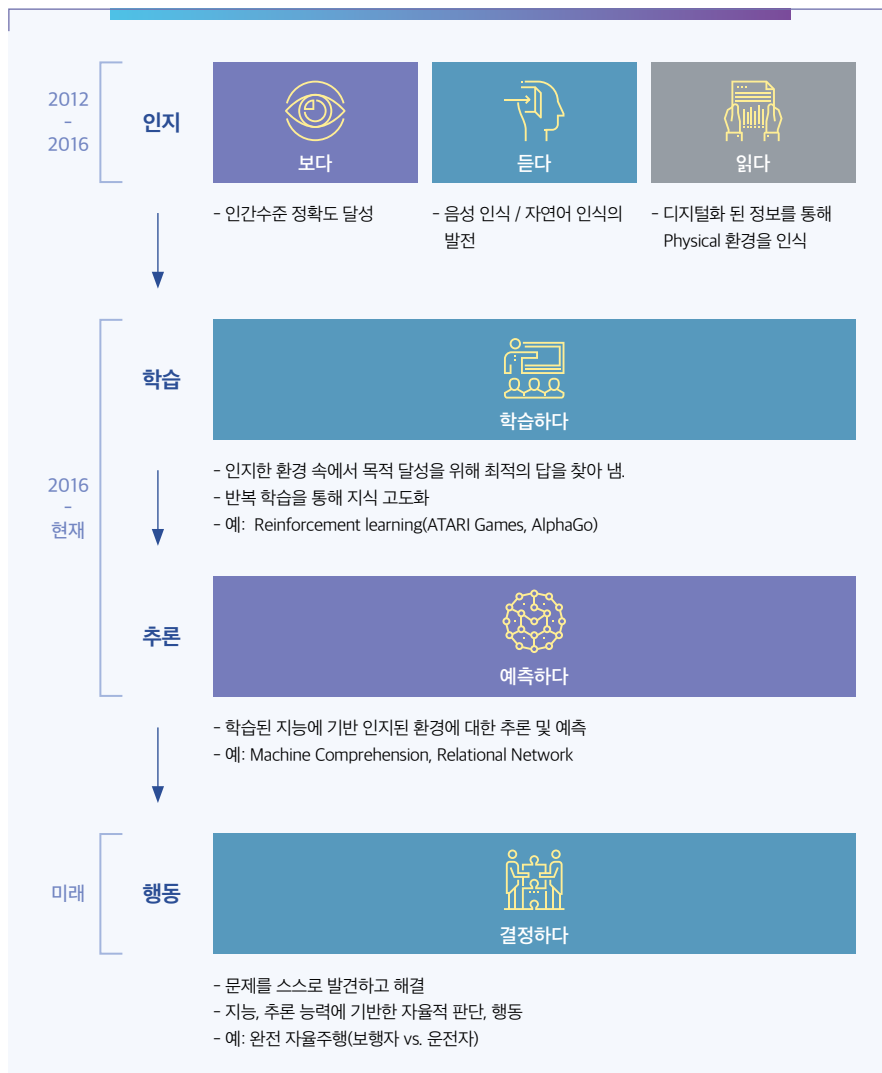
“우리에게 주어진 큰 기회다.”



- 손정의(손 마사요시) 소프트뱅크 회장

불과 수십 년 전만 해도 우리가 이처럼 일상에서 흔하게 사용하게 될 것이라고 쉽게 상상할 수 없었던 컴퓨터, 모바일 폰 등이 이제는 없어서는 안 될 물건이 된 것처럼 인공지능도 혁명적 변화의 일부며 지금이 새로운 기회를 잡을 수 있는 적기일지도 모른다. 이런 흐름 속에서 우리는 어떤 미래를 설계해야 하는지, 정책입안자는 사람들의 안녕한 미래를 위해 어떤 역할을 수행해야 하는지 등에 대해 깊은 고민이 필요하다. 본 보고서가 이러한 고민들을 해결해 가는데 조금이나마 도움이 될 수 있기를 희망한다.

◆ [그림 3] AI 기술의 발전 과정과 미래



출처 : 이승훈, "최근 인공지능 개발 트렌드와 미래의 진화방향", LG경제연구원, 2017. 12.

컴퓨터 비전·의약·비즈니스 분야의 머신러닝 개발과 적용 현황

1. 컴퓨터 비전과 머신러닝, 컴퓨터에게 보는 법을 가르치다

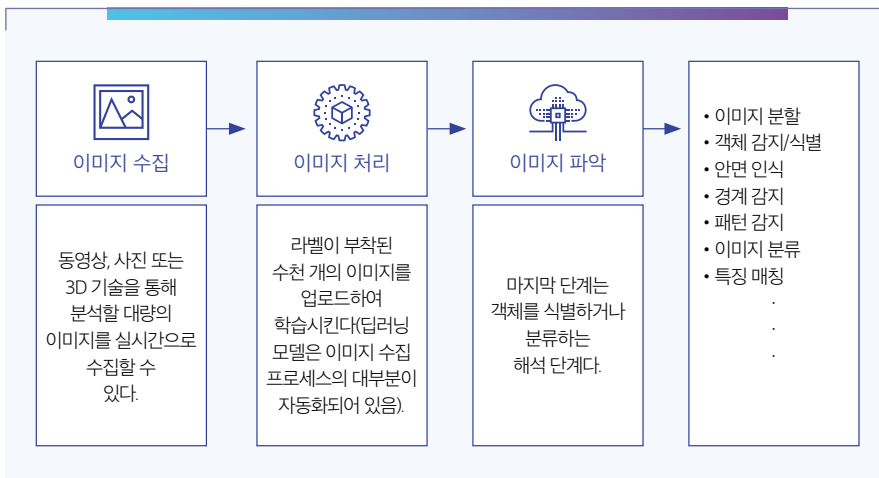
◆ 미래기술의 핵심, 컴퓨터 비전

“ 컴퓨터 비전(Computer vision)은 수학적 알고리즘을 통해 사람의 눈과 뇌가 하는 작업을 컴퓨터가 수행하도록 하는 것이라고 할 수 있죠. 인체 중에서 가장 모방하기 어렵다는 ‘눈’을 따라 해야 하기 때문에 끊임없는 연구와 도전이 필요한 분야이고, 최근에는 시각 정보뿐만 아니라 청각, 촉각 등의 감각분석까지 통합하는 추세로 발전하고 있습니다. 드론, 무인자동차, 스마트 기기 등이 급속도로 상용화가 진행되는 배경에는 인공지능, 특히 컴퓨터 비전 기술의 발전이 자리 잡고 있습니다. 컴퓨터 비전은 IoT와 무인자동차 시대에 새로운 “비전”을 제시할 것입니다. ”

예컨대 개와 고양이 사진이 있다고 가정할 때, 사람은 사진을 보고 개와 고양이를 쉽게 구별할 수 있지만, 컴퓨터는 그렇지 않다. 이때 컴퓨터가 고양이와 개의 사진을 계속해서 보고, 학습해서 개와 고양이를 구분해 낼 수 있도록 하는 기술이 머신러닝(기계학습)이다. 컴퓨터가 수많은 데이터를 학습해 사물을 군집하거나 분류하는 법을 익히도록 하는 것이 핵심이다. 이때 머신러닝으로 개와 고양이를 다른 객체로 인식하기에 앞서 이미지를 분류해주는 작업이 필요한데, 컴퓨터 비전이 바로 이 작업을 수행한다. 더 나아가 컴퓨터 비전은 컴퓨터가 세상을 3차원적으로 이해하도록 한다. 실제 고양이와 고양이 사진을 보여주면 컴퓨터는 3차원 정보 해석을 기반으로 한 평면에 들어있는 사진 속 고양이와 실제 고양이를 구분해낸다.

이렇게 컴퓨터 비전은 영상으로부터 의미 있는 정보, 특히 3차원적 정보를 추출하는 학문으로 사람의 눈과 뇌가 하는 작업을 수학적 알고리즘을 통해 컴퓨터가 수행하게 한다. 즉, 머신러닝이 ‘뇌’의 역할을 한다면 컴퓨터 비전은 인간의 ‘눈’에 해당되는 역할을 한다고 할 수 있다. 컴퓨터 비전 머신러닝 기술은 ▲이미지 속 사물을 구분하는 분류(Image Classification) ▲다양한 객체를 인식해 박스로 표시하는 검출(Object Detection) ▲대상의 형태와 경계를 나누는 분할(Image Segmentation) 등이 있다. 뿐만 아니라 AI로 이미지를 만들어 내는(Image Generation) 분야로까지 확장하여 다양한 분야에 적용되고 있다.

◆ [그림 4] 컴퓨터 비전의 프로세스



페이스북에 올린 사진에서 친구의 얼굴에 이름이 자동으로 입력되는 것이 머신러닝을 이용한 컴퓨터 비전의 대표적인 예라고 할 수 있다. 그리고 Flickr 같은 사진 공유 웹사이트에서 찾은 수 천 장의 로마 콜로세움 사진으로부터 콜로세움의 3차원 영상을 복구하는 것이 3차원 영상 재구성의 대표적인 예이다. 우리는 콜로세움을 찍은 두 장의 사진에서, 사진 1의 어느 점이 대략적으로 사진 2의 어느 점에 해당되는 지 쉽게 알 수 있지만, 컴퓨터는 정확한 대응점을 찾는 법을 가르쳐줘야만 한다. 컴퓨터 비전의 가장 큰 문제 중 하나인 이 대응점을 찾는 문제를 해결하기 위해 여러 가지 기법이 개발되었고, 3차원 영상 재구성의 속도와 정확도 향상을 위해 혹은 기존의 방법으로 대응점 검출이 어려운 경우를 위해 머신러닝을 이용하는 대응점 매칭 기법이 연구되고 있다.

머신러닝을 이용한 컴퓨터 비전은 다양한 분야에서 각광받고 있다. 애플이 야심차게 추진하고 있는 자율주행자동차 프로젝트 ‘타이탄(Titan)’의 핵심 중 하나가 바로 컴퓨터 비전 기술이다. 자율주행자동차는 카메라 등과 같은 시각 입력 장치를 통해 주변 세상을 인식하여 디지털 지도를 만든다. 이때 센서를 통하여 수집된 데이터를 이용하여 차선, 다른 차량, 보행자들을 찾아내는 머신러닝 알고리즘이 사용된다.

또한 지난 2015년부터 중국 공안이 범죄자 추적을 목표로 도입한 ‘텐왕(天網, 하늘의 그물)’의 경우 최대 1만 명의 얼굴을 비교해내는 컴퓨터 비전 안면인식 시스템을 통해 군중 속에서 검거 대상과 일반 시민을 구분해 내고 있다. 한국기계연구원의 ‘머신비전’의 경우, 기계장치가 정상작동할 때 발생하는 데이터를 입력하고, 이를 기반으로 머신러닝으로 훈련하도록 하여 기계장치의 고장이나 이상 증상 발생 시 수리가 필요한 위치를 정확히 제시한다.

인공지능 기술발전에 힘입어 최근 컴퓨터 비전 분야에서는 카메라를 통해 획득한 2차원 영상을 활용하여 3차원 정보를 정확하게 인지하는 기술에 대한 연구가 활발히 진행되고 있다. 4차 산업혁명에서 주목받고 있는 분야인 자율주행자동차, 가상/증강현실, 3D프린팅 정밀가공 등을 실현하기 위해서는 주변 환경과 대상 객체들에 대한 정확한 3차원 정보를 해석하는 것이 매우 중요하기 때문이다.



최근 미국에서는 퓨즈(Fyuse)라는 앱이 개발되었다. 이 앱은 컴퓨터 비전과 인공지능 기반 머신러닝을 활용하여 보통의 스마트폰에 탑재된 카메라로도 3D 이미지를 쉽게 구현할 수 있게 한다. 이 앱을 사용하면 사람이나 사물을 주위에서 찍은 후 3D 영상처럼 360도로 돌려보거나 자신 주위의 배경을 왜곡 없이 파노라마 형태로 볼 수 있다. 스마트폰에 탑재된 가속도계와 자이로스코프, 마그네틱 센서, 위치기반 정보 등을 융합한 기술의 결과물이다.

이와 같이 인공지능 기술을 활용해 3차원 정보를 인지하는 연구는 ▲스테레오 영상으로부터 머신러닝 및 딥러닝 기술을 이용하여 깊이를 추정하여 대상 객체의 3차원 포인트 클라우드를 획득하는 기술에 대한 연구 ▲대상 객체에서 3차원 정보를 복원하는 기술과 3차원 포인트 클라우드를 획득하는 기술에 대한 연구 ▲사람의 행동이나 동작에 대한 인식을 위해 2차원 영상으로부터 사람에 대한 3차원 자세를 추정하는 기술에 대한 연구 등이 활발하게 진행되고 있다.

3D 영상 재구성의 다양한 상용화를 위해 대상에 특정한 패턴이 있는 빛을 투영하여 인위적인 특징점을 발생시킨 후 촬영하는 (예를 들면 레이저 그리드를 피사체에 투영하여 인위적으로 특징점을 생성하고 실시간 분석을 거쳐) 3차원 영상을 재구성하는 방법이 개발되고 있다. 현재 모든 상용화된 3차원 스캐너는 이 능동적 재구성 방법을 사용하는데, 특징점의 즉각적 인식이 이뤄진다는 점에서 스테레오 영상으로부터 3차원 영상을 재구성하는 수동적인 방법에 비해 유리한 반면, 패턴을 투영해야 하기 때문에 대상과 비교적 가까이 있어야 한다는 근본적인 제약이 있다. 능동적 재구성 방식을 보안에 응용한 대표적인 예가 애플 아이폰의 페이스 아이디(Face ID)이다. 적외선 프로젝터가 대상을 향해 약 3만 개 이상의 랜덤 패턴 점을 쏘면 적외선 센서를 탑재한 카메라 렌즈가 이러한 패턴을 읽어내 원본과의 일치 여부를 가려내는데, 이때에도 사용자의 얼굴의 3차원적 특징을 학습하는 일종의 머신러닝이 주요하게 쓰이게 된다.

💡 컴퓨터 비전은 어떻게 미래기술의 핵심으로 도약했나?

“결과는 종종 혁신적이지만 진화는 항상 점진적이다”

- 기술의 진화: 비유와 함의들, 이관수(동국대 다르마칼리지 이관수 교수)에서 -

컴퓨터 비전은 1960년대 AI와 로보틱스 연구에서 출발했다. 로봇이 쌓여 있는 나무토막을 시각적으로 인지해 옮기는 실험을 통해 널리 알려졌으나, 현실 세계와 인간으로 대상을 넓혀 적용하자 컴퓨터가 인간처럼 상황을 인지하고 대상을 분석하는 것이 어려웠고 연구에서 상업화까지 이어질 만한 동인이 없었다. 이후 국제 컴퓨터비전 학회 등을 중심으로 기초연구가 꾸준히 진행되었고, 1990년대에 접어들며 테러 등 국가 보안 분야에서 얼굴 인식에 대한 니즈가 높아지면서 인지 중심의 컴퓨터 비전이 다시금 주목받기 시작했다. 2010년을 전후해서 스마트폰을 중심으로 대중적인 영상데이터가 폭발적으로 증가하였고 분석 대상이 되는 자료의 범주가 넓어지자, 이러한 자료들을 분석하는 기술이 중요해지면서 컴퓨터 비전이 차세대 기술 발전의 중심으로 자리하게 되었다. 나아가 사람과 비슷한 방식으로 세상을 경험하고 예측하며 행동을 결정하는 알고리즘이 4차 산업혁명을 주도하고 있다. 컴퓨터 비전과 관련된 유망 분야로는 자율주행시스템, 의료영상 보조진단시스템, 영상보안 및 검색시스템, 로봇의 시각인지시스템, 스마트 기기의 비서시스템 등이 대표적이다.

◆ 컴퓨터 비전과 머신러닝 기술을 구현하는 수학

머신러닝을 이용하여 더 심오한 진실을 발견하기 전에

“ 우리는 먼저 머신러닝에 관한 심오한 진실을 발견해야만 한다. ”

<마스터 알고리즘 - 머신러닝은 우리의 미래를 어떻게 바꾸는가 중>

앞으로 컴퓨터 비전 분야에서 머신러닝의 입지는 점점 더 커질 것이다. 해당분야 연구자들은 머신러닝이 현재 계속해서 개발 중인 기술인만큼, 완전한 단계에 들어섰다고 단언하긴 어렵지만, 잠재력과 한계 역시 단정 지을 수 없다고 말한다. 따라서 컴퓨터 비전 분야를 더욱 발전시키기 위해서는 머신러닝 및 딥러닝 알고리즘을 향상시키고 수학적 접근법을 이용한 방법과의 융합, 상호보완적인 연구를 계속해 나가려는 노력이 필요할 것이다.

사실 딥러닝의 두 가지 핵심 아이디어인 합성곱 신경망과 역전파는 이미 1989년에 소개되었다. 시계열을 위한 딥러닝의 기본인 LSTM(Long Short-Term Memory) 알고리즘은 1997년에 개발되었으며, 현재까지는 거의 변화 없이 이어져 왔다. 일반적으로 많은 전문가들은 하드웨어, 데이터, 알고리즘의 세 가지 기술적인 힘이 머신러닝의 진보를 이끌었다고 보고 있다. 2009년부터 2010년 사이에 몇 가지 간단하지만 중요한 알고리즘이 개선되면서 딥러닝이 빛을 발하기 시작했을 만큼 컴퓨터 비전 분야에서 알고리즘 개발은 부분이자 전체라고 할 수 있을 만큼 중요하다.

“ 수학자들은 신은 물리법칙을 어길 수 있지만 논리 법칙은 신이라도 거역할 수 없다고 말한다. ”

이렇게 머신러닝을 가능케 하는 알고리즘의 개발은 수학을 제외하고서는 설명할 수 없다. 머신러닝 이론은 통계학, 확률론, 해석학 등 수학의 다양한 분야와 컴퓨터 과학이 교차하는 분야이다. 머신러닝 및 딥러닝의 엄청난 가능성에도 불구하고 좋은 결과를 얻으려면 알고리즘의 내부 작동 원리와 논리적인 구조를 잘 파악해야 하며 많은 부분을 수학적으로 완전히 이해해야 한다.

머신러닝 알고리즘은 수학자들의 고민을 컴퓨터에 접목시킨 것이다. 수학은 머신러닝 탄생에 결정적 역할을 했으며, 동시에 머신러닝 발전의 열쇠가 될 수 있다. 18~19세기에 가우스와 오일러가 데이터 피팅 모델을 통해 선보인 머신러닝의 개념이 컴퓨팅 기술이 발전한 현대에 와서야 빛을 본 것처럼, 수학은 새로운 분야의 개척을 위한 전략이자 기회로 사용될 수 있을 것이다.

2. 머신러닝, 신약개발의 신(新)이 되다



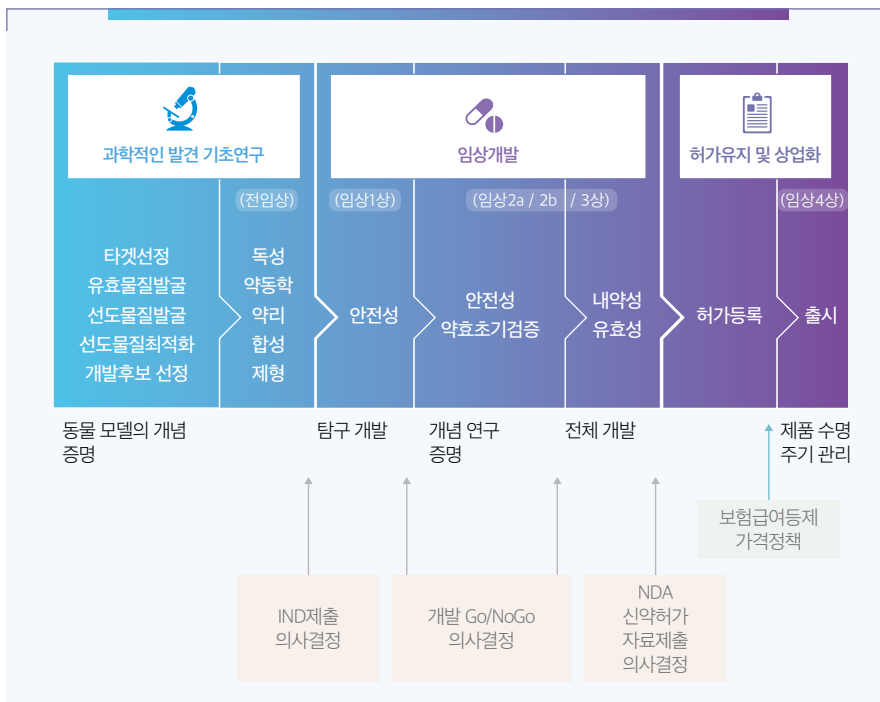
첫 번째
의제

◆ 의약연구의 패러다임을 바꿀 신기술의 등장

최초로 중국 대륙을 통일한 진나라 시황제는 먹으면 늙지 않는다는 불로초를 찾기 위해 중국 전역에 신하 수천 명을 보냈다고 한다. 진시황이 ‘신약’에 쏟아 부은 돈과 시간을 고려하면 21세기를 살아가는 우리 모두는 황제라고 할 수 있다. 약효를 인정받는 신약 한 개가 탄생하려면 무려 10~15년의 기간 동안 1~2조 원의 금액이 필요하기 때문이다. 모래사장에서 바늘 찾기만큼이나 어렵다고 표현되는 신약개발이기에 인공지능(머신러닝)에 거는 기대가 그 어느 분야보다도 크다.

이룸(Eroom)의 법칙이란 것이 있다. 이는 무어(Moore)의 영문 철자를 거꾸로 표기한 것으로 1950년 이후 제약회사에서 신약 개발에 투자하는 연구비 10억 달러당 미국 식품의약국(Food and Drug Administration, 이하 FDA)에서 승인하는 신약의 수가 9년마다 반으로 줄어드는 현상을 설명한다. 이는 신약개발이 오래 걸릴 뿐만 아니라 시간이 지남에 따라 더욱 많은 비용이 들어간다는 사실을 설명하고 있다.

◆ [그림 5] 신약 개발 파이프라인



신약개발은 △과학적인 발견 기초연구 △임상개발 △허가유지 및 상업화 등이 포함된 복잡다단하며 순차적인 개발단계를 거친다. 이를 통해 약물성(druggability)과 유효성(efficacy)이 엮보이는 수도 없이 많은 파이프라인(Pipeline) 중에서 단 한 개라도 승인을 받으면 성공이라 부른다. 더욱 많은 후보물질을 확보하면 투자 비용이 높아지겠지만 보다 많은 파이프라인을 확보하여 다각도의 개발 관점에서 연구할수록 신약개발 성공 확률이 높아진다. 신약개발은 성공을 한다면 매우 높은 수익을 창출하겠으나 낮은 성공 확률에도 불구하고 많은 시간과 자본을 투입해야 하는 고위험·고수익(High Risk, High Return) 산업이다.

신약개발의 전문학적인 연구개발비와 소요기간을 줄이기 위한 노력은 꾸준히 이뤄지고 있으며 빠르게 성장하는 머신러닝 기법과 기술도 하나의 대안으로 주목받고 있다. 연구자 한 명이 조사할 수 있는 자료가 한해 200~300여 건이라면, 머신러닝은 100만 건 이상의 논문을 확인하고 400만 명 이상의 임상데이터 분석이 가능하기 때문이다. 머신러닝을 통한 신약개발은 거스를 수 없는 세계적 추세이며, 이미 2016년부터 글로벌 제약사들은 기술력을 인정받은 인공지능업체들과 공동연구를 진행하고 있다. 화이자(Pfizer), 테바(TEVA) 등은 IBM의 인공지능 컴퓨터 왓슨(Watson)과 손을 잡았으며 일본은 2017년부터 국가적으로 인공지능 신약 개발에 돌입했다. 우리나라의 한미약품은 임상실험에 인공지능 플랫폼을 도입했고 CJ헬스케어, JW중외제약 등도 전 세계적인 흐름에 발맞추고 있다.

머신러닝과 신약개발의 만남은 제약업계의 의사결정 구조 및 협업의 방식에도 큰 영향을 끼치고 있다. 신약개발은 개발 성공률 대비 투입되는 R&D 비용이 워낙 큰 분야이기 때문에 조금이라도 위험 요소가 존재하거나 결과가 불확실하다고 판단되면 그 즉시 과정을 멈추고, 개발을 취소하는 것이 가장 효과적이고 일반적인 전략(신속의사결정모델)으로 여겨진다. 하지만 아무리 많은 지식과 경험을 지닌 유능한 의사결정자라 하더라도 신속한 결정을 내리기 위해 모든 사항을 고려하여 수치화하기 보다는 가중치를 두는 결과에 기반해 직관적 혹은 경험적 판단으로 결정하는 경우가 있다. 관련 전문가들은 머신러닝이 이러한 역할을 대신하여 많은 자료를 객관적으로 판단해줌으로써, 보다 빠르고 정확한 의사결정을 내릴 수 있을 것이라고 기대하고 있다.

또한 기존에는 연구위원급(고경력 연구자)-책임연구원급(박사)-연구원(석사 또는 학부생)의 수직적인 관계 속에서 더 많은 지식과 경험을 가진 사람이 의사결정을 주도하였다. 하지만 머신러닝이 도입되면서 알고리즘에 대한 해석 및 활용 능력의 중요성이 커졌으며, 최고 전문가의 판단 만큼이나 실무자 혹은 젊은 인재들의 생각과 판단 역시 중요해졌다. 학문과 기술개발의 관계와 패턴이 점차 변화하고 있는 만큼 연구자 간 협업의 방식 또한 혁신이 필요해진 시점이라고 할 수 있겠다.

◆ 신약개발 성패의 열쇠, 신뢰할 수 있는 데이터

신약개발 전주기에서 인공지능이 사용된다고 하지만, 인공지능이 신약을 만드는 건 아니다. 인공지능, 즉 머신러닝은 문제 해결을 위한 도구로서, 문제 해결을 위해 보다 적절한 실험 설계를 하거나 결과의 생성 및 예측을 모델링하는 과정에서 활용된다. 이에 무엇보다도 재현성이 있고 신뢰할 수 있는 데이터의 확보가 중요하다. 데이터의 양과 질에 따라 기계학습의 성공여부가 결정되기 때문이다. 하지만 다른 분야에 비해 바이오메디컬 분야의 데이터 생산 비용은 매우 비싸며(타 분야 평균비용 20불, 바이오메디컬 분야 평균비용 1만 불) 유의미한 정보의 비율은 적은편이다. 또한 이 데이터는 △사람의 생명(건강)과 관련된 정보인 점 △실패한 물질이더라도 향후 이 물질로 다른 연구를 진행해 좋은 결과를 낼 수도 있다는 점에서 굉장히 소극적으로 공유되고 있다.

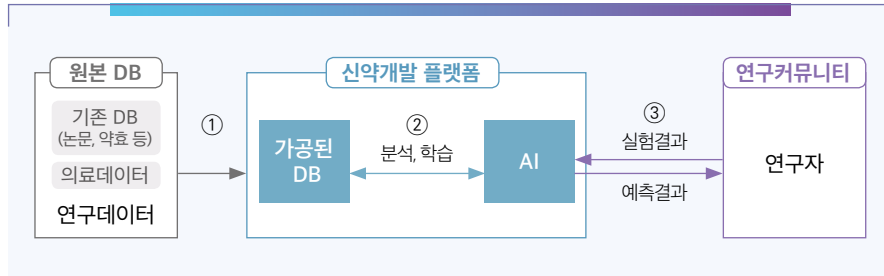
현재 국내 헬스케어 데이터는 개인정보보호법 등에 따른 규제로 인해 공유가 어려운 실정이다. 국내 연구진은 먼저 미국 FDA 공개 데이터를 포함해 세계적 의약 연구데이터부터 학습시키고 있다. 한국인을 위한 신약을 개발하고 있으면서 해외에서 수집·작성한 데이터나 알고리즘을 공유 받아 사용하는 것은 안타까운 일이다. 환자의 권익을 침해하지 않고 합법적으로 신약개발에 필요한 데이터를 사용할 수 있는 구체적인 대안이 마련돼야 할 것이다. 이를 위해 연구목적으로의 데이터 공유는 실패한 결과물을 포함한 모든 데이터에 대해서 수집·활용하는 것을 허가하는 방향으로 제도를 개선할 필요가 있다.



● 신약개발에 활용되는 데이터

- 기존 신약 연구자 주도 분자 구조 및 약리/물성/독성 데이터 (신약 허가에 필요한 직접적 데이터)
- 시스템 생물학 연구자가 주도하는 맞춤형 의약품 개발을 견인하는 옴릭스 데이터
- 의약품 재심사/재평가 내역, 의료 보험, 인체자원은행 등 국가 주도 보건의료 데이터
- 병원 주도 헬스케어 데이터 (의무기록, 진단, 검사, 처치, 처방)
- 환자 주도의 SNS 등에 올린 셀프 보고서 (ex. 약물 순응도, 비용 대비 만족도)
- 제약사 및 협력사가 만든 신약 파이프라인의 경제성 평가 및 마케팅 자료
- 데이터 사용을 위한 인프라 (모바일 헬스 어플) 제공자에 의해 가공 및 정리된 2차 데이터 등

◆ [그림 6] 연구데이터 공유·활용 추진계획(안)



이에 앞서 산·학·연 공동연구를 통해 각 기관들이 자신들의 데이터를 공동연구기관들과 공유할 수 있도록 함으로써 양질의 데이터를 보다 수월하게 대량으로 확보할 수 있도록 해야 한다. 과학기술정보통신부와 보건복지부는 올해 ‘빅데이터·인공지능 기반 신약개발 플랫폼 구축’ 사업방안을 발표하며 ‘국가연구개발사업을 통해 생산된 50만 건의 화합물 연구데이터에 인공지능을 적용해 신약개발 시간과 비용을 1/3로 단축’하고자 한다고 밝혔다. 이 플랫폼을 활용하여 제약사들도 지적재산권에 크게 저촉되지 않는 수준과 내용의 데이터를 공개할 수 있게 된다면 긍정적인 시너지 효과를 낼 수 있을 것이다.

원론적으로 말하자면 데이터의 양적, 질적 우수성이 예측 모델의 예측 범위와 정확성을 결정하는 가장 중요한 요소이므로 이러한 플랫폼을 통해 향후 확보되는 국내외 빅데이터는 머신러닝 기반 예측모델의 눈부신 발전의 근간이 될 것이다. 이와 더불어 전문가들은 AI 신약개발이 성공하기 위해서는 ▲공동연구 증가 및 다각화 시도 ▲인력교류 및 양성 ▲바이오메디컬 데이터 공유 활성화 ▲활용분야의 다양화 ▲결과 검증 강화 ▲성능 향상 등의 노력이 필요하다고 입을 모은다.

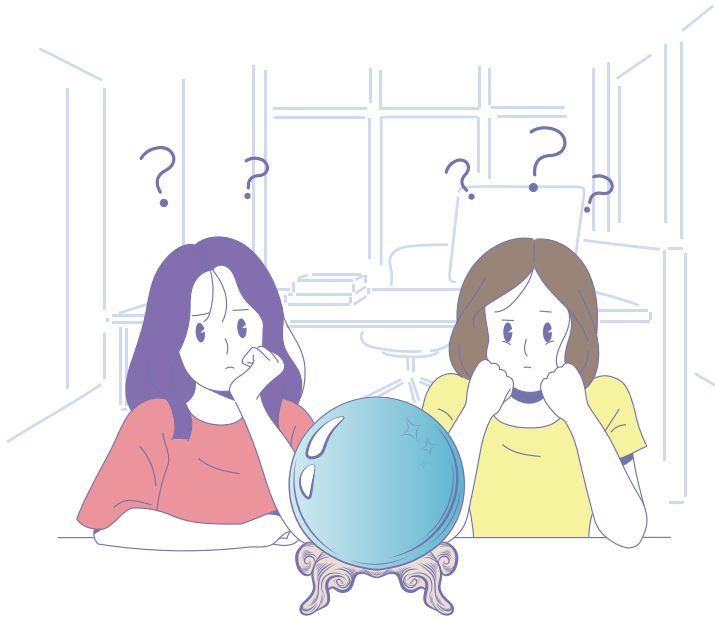


3. 미래를 향하는 나침반, 비즈니스 인텔리전스



첫 번째
의제

◆ AI와 BI(비즈니스 인텔리전스)가 만났을 때



미래의 불확실성에 기인한 궁금증과 불안은 인류의 뿌리 깊은 고민거리이다. 연말연시가 되면 ‘점집’은 문전성시를 이루고 무료 사주, 토정비결 서비스를 제공하는 애플리케이션은 다운로드 순위에서 늘 상위권을 기록한다. 개개인 차원에서도 미래에 대한 궁금증을 해결하기 위해 이러한 노력을 기울이는 것을 볼 때 국가나 사회 전반에 영향을 주는 이슈와 관련된 미래를 예측함에 있어서는 더욱더 정확한 데이터를 기반으로 예측해야 하는 영역으로 넘어가야 함을 알 수 있다.

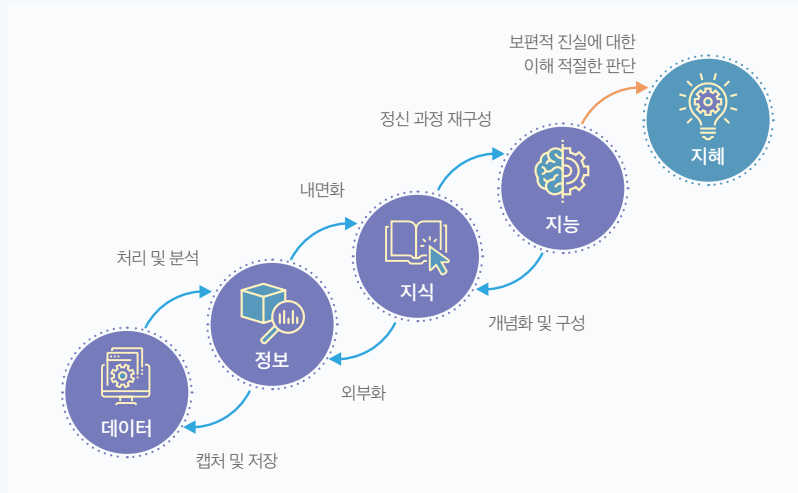
하지만 미래는 예측하기 굉장히 어렵다. 그래서 아직도 수많은 사람들이 미래를 알기 위해 끊임없이 고민하고 있다. 미래에 영향을 미치는 요인들은 매우 많고, 그 요인들은 서로 다른 시간차를 두고 영향을 끼치기 때문이다. 이에 미래 우리의 산업, 나아가 국가에 영향을 줄 수 있는 기술적 기회와 위협 요인을 발굴해 적절한 시점에, 알맞은 형태로, 적합한 의사결정자에게 배포할 수 있도록 지원하는 모든 방법론을 다루는 기술 인텔리전스의 중요성이 날이 갈수록 커지고 있다.

기술 인텔리전스는 미래 비즈니스에 영향을 줄 수 있는 것으로 예측되는 정보를 수집, 분석, 시각화하여 최적의 의사결정을 내릴 수 있도록 지원하는 프로세스인 비즈니스 인텔리전스 중 하나이며, 여기에는 고객과 관련된 정보를 다루는 고객 인텔리전스, 경쟁자에 관련된 정보를 다루는 경쟁 인텔리전스 등도 포함된다.



지혜가 불변의 진리를 의미한다면 지식과 지혜 중간 단계에 있는 것이 인텔리전스다. 단계별 예시로 보자면,

- 데이터(Data): 가공되지 않은 단순한 형태 (분야별 일일 판매량과 마케팅 활동)
- 정보(Information): 데이터를 분석가공한 형태(분야별 월별 판매량 추이)
- 지식(Knowledge): 정보로부터 의미를 찾아낸 형태(마케팅 활동이 분야별 판매량 증감에 미치는 영향)
- 인텔리전스(Intelligence): 지식을 재구성해 결론을 도출해낸 형태(새로운 마케팅 활동이 분야별 판매량 증감에 미칠 것으로 기대되는 영향)
- 지혜(Wisdom): 불변의 진리



비즈니스 인텔리전스에서 가장 중요한 요소는 정보의 수집과 분석이다. 컴퓨터가 수많은 데이터를 통해 스스로 패턴을 찾아내고 이를 분류나 예측에 적용하는 알고리즘인 머신러닝이 적극적으로 활용되고 있으며, 그간 이루어진 방법론에 큰 영향을 미치고 있다. 기존에는 대부분의 자료가 숫자와 같은 구조화된 데이터이거나 비구조화된 데이터 중에서도 텍스트 형태의 것이었다면, 머신러닝은 음성, 동영상, 이미지 등 다양한 형태의 자료를 객관적으로 빠르게 수집할 수 있도록 지원해주고 있다. 고객의 상품 리뷰가 담긴 음성 파일, 경쟁사의 소개 영상 등은 중요한 데이터이지만 해당 데이터를 사람이 다루는 순간 주관의 개입 위험이 존재하여 활용의 어려움이 있었다. 하지만 머신러닝의 발전은 주관 개입의 개연성을 최소화함으로써 활용 가능한 자료의 범위를 극대화 시켜주었다.

... 많은 이들이 머신러닝 모델을 활용해 미래를 '예측'한다고 말한다. 머신러닝 모델은 과거 경험(데이터)로부터 특정한 규칙(패턴)을 추론하는 메커니즘을 가지고 있다. 이에 따라 머신러닝의 예측은 '미래에 어떠한 입력 값들이 주어졌을 때 우리가 관심 있는 값은 어떨 것이다' 라고 유추하는 것을 의미한다.

또한 데이터 분석 방법에 있어서도 큰 영향을 끼치고 있다. 수많은 요인에 대한 완벽한 통제가 불가능한 우리 사회의 자연적 조건 속에서 사람이 통찰력을 가지고 미래를 예측하기란 쉽지 않다. 이때 머신러닝은 수집된 방대한 양의 데이터를 토대로 패턴을 도출해낸다. 예를 들어, 미래 유망특허를 발굴한다고 할 때 기존에는 미래에 유망할 것으로 예상되는 기술의 특허가 가져야 하는 몇 가지 특징을 설정하고 이에 해당하는 특허를 추출해 내는 방법을 사용했다. 하지만 머신러닝은 과거에 유망했던 특허들을 빠르게 분석하고 공통된 패턴을 찾아낸 뒤 유사한 특징을 갖는 특허를 발굴해낸다. 이는 사람이 미처 포착해내지 못했던 부분까지도 도출해낸다는 점에서 유용하다.

성과 분석의 영역에서 주로 활용되던 비즈니스 인텔리전스는 머신러닝과 결합되면서 예측의 영역까지 그 활용이 확대되었다. 최근까지만 해도 알고리즘 코드를 작성하려면 데이터 전문가가 필요했으나, 지금은 몇몇 데이터 포인트를 클릭하고 예측하고자 하는 변수를 선택하기만 하면 예측모델이 자동으로 생성되기 때문에 마케터들이 고객 행동을 예측 및 대응하는 영역에 적극 활용하고 있다. 또한 비즈니스 관리자들은 위험 점검 및 예측에, 공급망 관계자들은 물류 점검 및 최적화 작업에 이를 활용하고 있다. 미국 정보기술연구 주식회사인 가트너의 데이비드 설리 부사장은 2018년 디지털 트렌드를 발표하면서 향후 몇 년 안에 비즈니스 인텔리전스의 90%가 인공지능을 접목시킬 것이며, 전문 데이터 과학자가 아닌 일반 직원이 활용할 수 있게 될 것이라고 했다. 인공지능이 사람을 대체할 것이라는 시각이 아니라, 인공지능은 사람의 의사결정 과정을 지원함으로써 미래를 내다보는데 도움을 주는 존재가 될 수 있다고 보는 것이다.

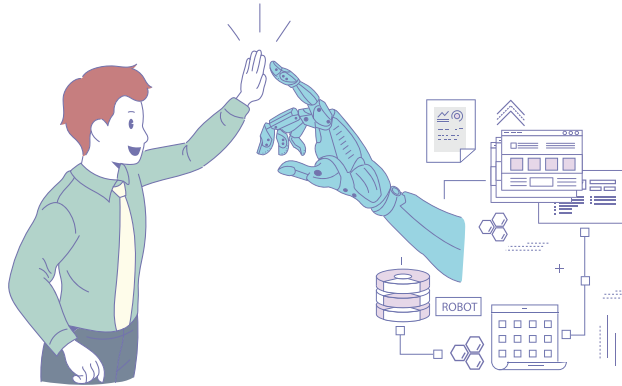


연구현장 Real Story

‘ㅇㄱㄹㅇ’, ‘우디르급 태세전환’, ‘인지용’... 일상생활에서는 전혀 접하지 않는 용어들인 만큼 게임에서 텍스트마이닝* 분석을 진행하다 마주하면 어리둥절할 때가 많습니다. 그래서 이렇게 저렇게 공부하게 되는 경우가 많은데 아직 많이 부족하고 저와 같이 생각하는 전문가들이 많은 만큼 급식체를 일반체로 전환하는 용어사전이 필요하지 않을까 생각합니다.

*텍스트마이닝: 반정형/비정형 텍스트 데이터를 정형화하고, 특징을 추출하며 추출된 특징으로부터 의미 있는 정보를 발견할 수 있도록 하는 기술이다.

◆ AI와 인간의 협업이 비즈니스 인텔리전스를 완성시킨다



머신러닝이 더 많은 영역에서 더욱더 많은 일들을 효율적으로 해내갈수록 우리 인간이 설 자리를 잃어버리게 되지 않을까 하는 우려가 있을 수 있다. 하지만 머신러닝은 분명한 한계와 위험을 지니고 있으며 이를 통제하고 해결할 수 있는 것은 우리 인간뿐이다.

우선, 우리가 바라는 바람직한 방향으로 비즈니스 인텔리전스를 완성시켜가기 위해서는 분석을 위해 어떤 데이터를 사용할 것인지에 대한 결정을 내려야 한다. 잘못된 입력(input)은 잘못된 결과(output)를 가져올 수 있기 때문에 어떤 데이터를 선택할지 결정하는 과정은 매우 중요하며 깊은 고민 없이 데이터 기반의 접근법만을 신뢰한다면 잘못된 의사결정을 야기할 수 있다. 예를 들어, 기술동향을 분석할 때 미국 특허청의 데이터를 활용한 결과와 유럽 특허청의 데이터를 활용한 결과는 매우 다를 수밖에 없는데 이를 간과하고 전자의 결론을 토대로 후자의 문제를 해결하고자 한다면 잘못된 판단을 내리게 될 가능성이 크다.

다음으로, 머신러닝이 미래를 예측할 수 없는 분야가 존재함을 인지해야 한다. 미래 예측에는 크게 두 가지가 있다. 하나는 현재에서 미래를 보는 “forecasting”이고 다른 하나는 바람직한 미래의 모습을 그리고 그것을 위해 현재에 무엇을 해야 하는지 찾는 “foresight”다. 머신러닝이 다루는 분야는 전자(forecasting)로, 과거의 패턴으로부터 미래를 예측한다. 하지만 “미래를 예측하는 가장 좋은 방법은 미래를 창조하는 것이다(The best way to predict the future is to create it).”라는 피터 드러커(Peter Drucker)의 주장처럼 우리 사회가 어떤 미래를 만들어갈지에 대해 사회적 공감대를 형성하고 이를 위해 현재 무엇을 할지 결정하는 후자(foresight)의 영역도 매우 중요하다. 이를 위해 우리 스스로가 수많은 논의를 거쳐 도덕적 가치관을 확립하고 이를 토대로 기계에게 정답지 데이터를 만들어주어야 한다. 자율주행차가 운전자와 보행자 중 한 명을 선택해야 하는 상황에서 어떤 결정을 내리도록 답안지를 줄 것인지는 인간만이 할 수 있고, 인간이 해야 하는 영역이기 때문이다.

머신러닝과 함께 나아가야 할 길



두 번째
의제

1. 우수 인재 확보에 머신러닝의 내일이 달려있다

“ IT강국 대한민국의 AI성적표는 낙제점 ”

대한민국은 'IT 강국'으로 통했다. 초고속 인터넷 보급률 1위, 부호분할 다중접속(CDMA) 방식 이동통신 상용화(1996년), 컬러 액정 휴대폰(2001년), 와이브로(2004년) 등 IT기술 분야에서 '세계 최초' 타이틀이 즐비하다. 올해는 세계 최초로 5세대(5G) 이동통신 서비스와 스마트폰 상용화를 시작했다. 그러나 전문가들 사이에선 '한국은 더 이상 IT강국이 아니다'라는 이야기가 나오고 있다. 가장 큰 원인으로 인재 부족 문제가 있다.

◆ [표 1] 국가별 AI 인력 분포 현황(2017년 기준)



	국가	Top AI 인재 수(명)	전체 AI 인재 수(명)	Top/전체 AI 인재 비율(%)
	미국	5,158	28,536	18.1
	중국	977	18,232	5.4
	영국	1,117	7,998	14.7
	프랑스	1,056	6,395	16.5
	캐나다	606	4,228	14.3
	일본	651	3,117	20.9

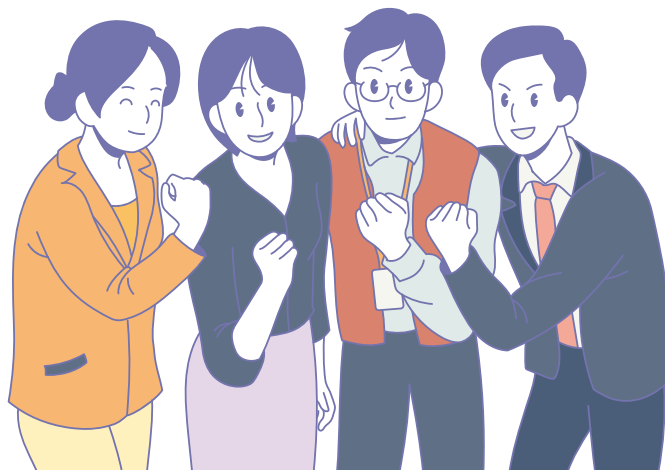
출처 : China Institute for Science and Technology Policy at Tsinghua University(2018)

미국 경제지 포브스(Forbes)는 지난해 한국이 AI 부문의 리더가 될 수 있는가를 주제로 한 기획기사를 통해 “한국은 인재가 부족해 중국을 따라가기 힘들다”고 평가했다. 우리나라의 AI 전문 인력은 양과 질 측면에서 터무니없이 부족한 실정이다. 2018년 중국 칭화대가 발표한 인공지능 보고서에 따르면 한국은 2,664명으로 주요 국가 15개 중 15위다. 미국의 AI 전문가는 우리나라의 10.7배, 중국은 6.8배, 일본은 1.2배 수준이며, 특히 연구 능력 상위 10% 이내 톱 전문가는 미국이 5,158명, 중국 977명, 일본 651명인 반면 우리는 이와 관련한 데이터조차 제대로 없어 0명으로 기록되어 있다.

4차 산업혁명 시대에는 막강한 힘을 가진 AI 기술과 인재의 보유 여부가 개인·기업·사회와 국가의 경쟁력을 결정짓고 지속 가능성을 좌우하게 될 것입니다. 우리나라가 글로벌 AI 주도권 경쟁에서

“앞서지 못하는 데에는 인재부족 문제가 가장 큰 원인이지만, 어떻게 인재를 육성해야 하느냐만 고민해야 할 것은 아닙니다. 국내 대학은 MIT로 갈 인재를 양성하는 인큐베이터라고 불릴 만큼 두뇌유출 문제도 심각합니다.”

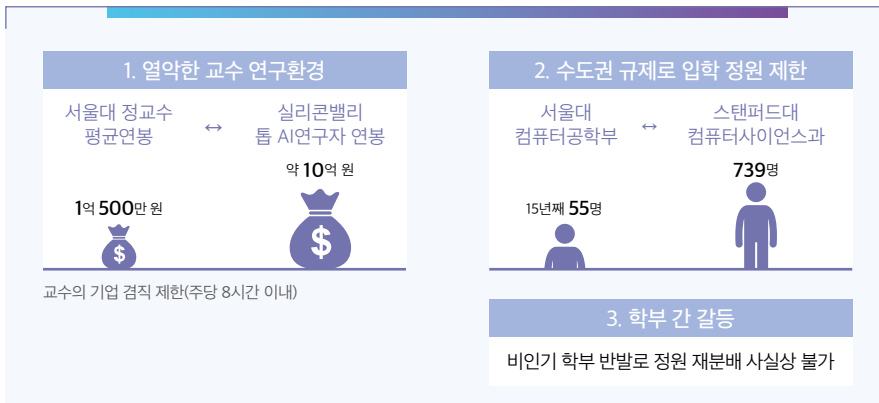
미국 UC버클리와 스탠퍼드대가 컴퓨터 전공자를 각각 1,590명, 739명으로 늘리는 동안 서울대 컴퓨터공학부는 15년째 제자리에 머물러있다(55명). 인재풀 자체가 작는데 그마저도 해외로 다 빠져나가면 국내 인재난은 더해질 수밖에 없다. 현재 전공을 불문하고 AI 기초 학문인 컴퓨터를 배우려는 학생들의 수요가 크지만 수도권 대학 정원 규제 때문에 정원 확대가 불가능한 실정이다. 신설되는 AI 대학원도 크게 다르지 않다. 한정된 정원으로 입학 경쟁률은 5~7대 1에 달하고, 예산 및 인력부족 등의 문제로 AI 전문가를 신규로 채용하기 보다는 기존 교수진으로 대체하여 커리큘럼을 꾸렸다. 안타깝게도 국내 대학은 학칙과 정원(定員) 규제, 예산 부족 등을 이유로 과거의 교육 체제를 그대로 답습하고 있다.



	카이스트	고려대	성균관대
정원	60명 (석사·박사)	50명 (석·박 통합, 박사)	60명 (석사·박사)
교수진	7명	7명	15명
AI전문가 신규 채용	없음	없음	없음
전임 교수의 전(前)소속 학부	전기전자공학(4) 전산(2) 산업시스템공학(1)	뇌공학(4) 컴퓨터(2) 바이오회공학(1)	SW(8) 전자전기공학(4) 반도체시스템공학(3)

출처 : 과학기술정보통신부, 각 대학

◆ [그림 8] 국내 AI 인재 육성의 주요 난제



출처 : '9월 AI 대학원 3곳 문 여는데, AI전문가 채용 없이 교수 돌려막기', 조선일보, 2019.05.04.

AI 인재의 유출을 막기 위해서는 인재들이 미래를 만들어 갈 수 있는 환경 조성이 필요하다. 미국 MIT(매사추세츠공과대학)는 올해 9월부터 이공계는 물론 인문·사회계열 등 전교생에게 AI 교육을 실시하고 교수 50명, 박사급 연구인력 100명 이상으로 구성된 AI 칼리지를 설립하여 연구와 더불어 교육과정을 개발하게 할 계획이다. 이와 대조적으로 국내 교육은 여전히 문·이과 분리와 학과별 칸막이 교육에 갇혀 있다. 이를 타개하고 적극적인 AI 교육을 확대할 수 있는 방안을 강구해야 한다. AI교육 확대를 위해 대학은 'AI+X' 복합 전공 등의 전공과목 추가 개설 및 교과목 신설 등 학과목 체계를 개선해야 하며, 고급 연구인력 양성을 위한 AI 석·박사 과정을 신설, 확대하고 장학금 및 해외연수 지원, 인턴십 확대 등 지원을 강화해야 한다. 또한, 우수 인재가 도전적이고 혁신적인 연구를 수행할 수 있도록 연구 여건을 조성하고, 글로벌 수준의 대우와 지원을 통해 AI 인재유출을 막기 위한 노력을 해나가야 한다. 그리고 우수 인재가 국내에서 활동할 수 있도록 스타트업 창업 등 산업 육성 지원과 규제 개선을 통해 건강한 AI 생태계를 조성하는 것이 중요할 것이다.

2. 차세대 과학자, 정책을 이야기하다

“ 어린 시절부터 컴퓨터를 자유자재로 다뤄본 사람은 신기술을 거부감 없이 받아들이는 방식에 익숙해지기 마련입니다. 이 같은 활용 마인드를 갖추지 않은 채 도메인 지식에만 골몰하는 교육은 지양해야 합니다. 즉, AI와 머신러닝에서 괄목할만한 성과를 이룰 인재를 배출하려면 수학에서 알고리즘의 이치를 깨닫고 공학에서 응용 방식을 배울 수 있도록 양쪽을 오가는 융합을 추진할 필요가 있습니다. ”

① 수학적 소양의 토대 위에 코딩과 프로그래밍을 가르쳐야 한다

최근 초등학교 정규 교과 과목에 소프트웨어 교육을 포함하면서 필수적으로 코딩을 배워야 하는 시대에 접어들었다. 유년기부터 컴퓨터를 자유자재로 다뤄보는 머신러닝이 대표하는 신기술 습득에 비교적 시간이 덜 걸리고 금세 적용할 수 있는 것이 사실이다. 다만 AI와 머신러닝에서 뛰어난 역량을 발휘할 인재를 양성하려면 공학적 응용력뿐만 아니라 수학에 대한 지식과 논리적 사고까지 갖추 수 있도록 교육해야 한다. 수학은 AI 알고리즘을 이해하는 기본 토대로, 머신러닝의 문제점을 찾아내고 해결하는 열쇠다. 일찍이 AI 기술을 선도한 선진국들이 대학 이전 단계의 교육과정에서 수리, 통계, 데이터 과학 등의 기초 소양을 쌓을 수 있도록 노력을 기울이고 있는 데는 그만한 이유가 있다. 고급 AI 인력을 단기간에 양성하는 것은 어려우며 수학, 컴퓨터 공학 등 기초과학과 연관 학문의 뒷받침이 전제되어야 하므로 초기 교육단계에서부터 AI 및 STEM 교육 투자를 통해 데이터 과학 및 컴퓨팅적 사고방식과 응용능력을 길러주어야 한다. 컴퓨터 과학의 대가 페드로 도밍고스 교수 역시 코딩을 가르치기 전에 수학과 논리를 교육한다면 훨씬 효과가 크다고 주장한 바 있다.

② AI 인력양성을 위해 산업계와의 협력이 중요하다

각 분야 기업의 AI·머신러닝 전문가 수요는 높지만, 공급은 한정적이기에 아쉬운 점이 많다. 따라서 산·학·관 간 적극적 협력을 통해 실무 능력을 갖춘 AI 인재 양성에 나서야 한다. 기업은 데이터의 수집부터 의사결정에 이르는 AI 개발의 전 과정을 포괄적으로 다룰 수 있어 업무 현장에 즉각 투입할 수 있는 실용적인 인재를 필요로 한다. 따라서 석·박사급 고급 인재 육성은 물론 대학교 학부 내 AI 교육 체계 신설, 기존 종사자의 재교육 프로그램 진행, 산학협력을 위한 공동 연구, 인턴십 등을 추진할 필요가 있다. 이때 기업과의 적극적 협력을 기반으로 교육 개발, 장학금, 연구비, 채용 등의 지원이 이뤄져야 한다.

미국, 유럽 등은 이미 국가가 나서서 AI와 머신러닝 분야의 전문가 양성과 함께 전략적 연구 지원 사업을 운영하고 있어요. 유럽 호라이즌 2020(Horizon 2020)의 빅 켐(Big Chem)이 그 대표적인 예입니다. 반면 우리나라는 주요 기관 중심으로만 프로젝트가 이뤄지고 있어 데이터 공유와 융합연구가 추진되기 어려운 상황이지요. 결국 국가 단위로 지원하는 다기관 융합 연구 프로젝트가 매우 절실합니다.

③ 다양한 분야 전문가와 머신러닝 전문가가 공동연구를 추진할 수 있는 장을 마련해야 한다

유럽 연합(EU)의 호라이즌 2020(Horizon 2020)은 미래 먹거리 사업 발굴을 위해 자본을 아낌없이 투자하는 거대 프로젝트다. 여기에 속한 빅 켐(Big Chem)은 신약 개발을 위해 머신러닝과 빅 데이터를 다루는 전문 인력을 육성한다. EU와 미국 국립보건원(National Institutes of Health, NIH) 등의 편당을 받은 연구 프로젝트에서는 각종 챌린지 대회를 열어 데이터를 공유하는 대신 각종 분석 아이디어와 기법을 수집한다. 우리나라도 도메인 지식과 머신러닝 기술이 협업을 할 수 있는 환경을 조성할 필요가 있다. 먼저 정부는 ▲기계학습에 필수적인 학습·평가 데이터와 컴퓨팅 인프라를 구축하고, ▲머신러닝 경진대회 등 기술 콘테스트를 개최해 데이터와 알고리즘을 공유하도록 하여 ▲각 분야에 활용하거나 공동연구를 추진할 수 있는 기회의 장을 열어줘야 한다. 나아가 이를 통해 공동연구를 추진하는 등 연구 역량을 결집하고자 하는 연구자에게는 지속적으로 연구개발비를 지원하는 방식으로 적극적인 R&D정책을 펴나가야 한다.

④ 데이터를 수집·확보·공유할 수 있는 인프라를 구축해야 한다

우수한 기계학습은 완벽한 알고리즘만으로 완성되는 것이 아니다. 양질의 데이터를 수집하고 학습할 수 있어야 이를 바탕으로 더욱 명확한 예측을 제시할 수 있다. AI 선도국들은 공통적으로 다양한 데이터셋이 구축되어 연구에 사용되고 있으나 국내에는 머신러닝 연구를 위해 활용할 수 있는 우리나라만의 고유 데이터셋이 부족한 실정이다. 따라서 데이터 수집·공유를 위한 인프라를 구축하고, 기 존재하는 데이터에 대해서는 연구 목적 사용이 가능하도록 규제를 완화하고, 이를 바탕으로 일군 일련의 성과도 공유하게 함으로써 다양한 응용연구에 활용할 수 있도록 해야 할 것이다. AI와 머신러닝이 헬스케어를 비롯한 타 산업에 적극 활용될 수 있도록 하기 위해서는 AI 활성화와 함께 데이터의 안전하고 합법적인 수집·활용을 위한 제도와 체계 개선의 노력이 필요하다.

참고문헌



- °과학기술정보통신부(2018), I-Korea 4.0 실현을 위한 인공지능(AI) R&D 전략
- °소프트웨어정책연구소(2016), 기계학습의 발전 동향, 산업화 사례 및 활성화 정책 방향
 - 딥러닝 기술을 중심으로, SPRI Issue Report 2015-017호
- °정보통신기술진흥센터(2017), AI 인재부족현상우려 국가별 대책마련에 분주, ICT Brief 2017-47
- °정보통신기획평가원(2016), 컴퓨터 비전과 인공지능
- °페드로 도밍고스(2016), “마스터 알고리즘 - 머신러닝은 우리의 미래를 어떻게 바꾸는 가”, 비즈니스북스
- °한국보건산업진흥원(2019), 주요 국가별 인공지능(AI) 인력양성 정책 및 시사점, 보건산업브리프 vol.276

한국과학기술한림원은,

과학기술 분야 한국을 대표하는 석학단체로서
1994년 설립되었습니다.

1,000여 명의 각 분야 연구리더들이 한림원의
회원이며, 각자의 역량과 지혜를 결집하여 기초
과학진흥을 위해 뛰고 있습니다. 국회와 정부
등 국가정책기관에 전문가 의견을 제시하고,
과학기술 분야 국제교류와 민간외교 활성화를
위해 노력 중이며, 국민들에게 한 발 더 다가가는
기관이 되기 위해 고민하고 있습니다.

한림석학정책연구는,

우리나라의 중장기적인 과학기술정책과 주요
현안에 대해 한림원 회원들이 직접 참여하는
연구 프로젝트로서 각 분야 전문가들의 지식과
의견을 담고 있습니다.

주제 및 보고서의 특징에 따라 한림연구보고서
(중·장기 비전 마련을 위한 정책연구 보고서),
차세대리포트(차세대회원들의 의견과 아이디어를
담은 정책제안서) 등으로 다채롭게 발간하고
있으며, 국회, 정부 등 정책 수요자들에게 꼭
필요한 지식을 제공하기 위해 꾸준히 노력하고
있습니다.

한림원에 대해 더 자세한 내용보기

홈페이지
www.kast.or.kr

블로그
kast.tistory.com

포스트
post.naver.com/kast1994

페이스북
www.facebook.com/kastnews



이 사업은 복권기금 및 과학기술진흥기금 지원을 통한 사업
으로 우리나라의 사회적 가치 증진에 기여하고 있습니다.



KAST 한국과학기술원
The Korean Academy of Science and Technology

(13630) 경기도 성남시 분당구 돌마로 42

Tel 031-726-7900 **Fax** 031-726-7909 **E-mail** kast@kast.or.kr

