KAKAO AI REPORT

Vol.01

2017.03

```python
import kakao.ai.dataset.daisy
import kakao.ai.image
import kakao.ai.classifier

import mxnet as mx

def Conv(data, num_filter, kernel=(1, 1), stride=(1, 1), pad=(0, 0), name=None, suffix=''):
    conv = mx.sym.Convolution(data=data, num_filter=num_filter, kernel=kernel, stride=stride, pad=pad, no_bias=True, name='%s%s_conv2d' %(name, suffix))
    bn = mx.sym.BatchNorm(data=conv, name='%s%s_batchnorm' %(name, suffix), fix_gamma=True)
    act = mx.sym.Activation(data=bn, act_type='relu', name='%s%s_relu' %(name, suffix))
    return act

def Inception7A(data, num_1x1, num_3x3_red, num_3x3_1, num_3x3_2, num_5x5_red, num_5x5, pool, proj, name):
    tower_1x1 = Conv(data, num_1x1, name=('%s_conv' % name))
    tower_5x5 = Conv(data, num_5x5_red, name=('%s_tower' % name), suffix='_conv')
    tower_5x5 = Conv(tower_5x5, num_5x5, kernel=(5, 5), pad=(2, 2), name=('%s_tower' % name), suffix='_conv_1')
    tower_3x3 = Conv(data, num_3x3_red, name=('%s_tower_1' % name), suffix='_conv')
    tower_3x3 = Conv(tower_3x3, num_3x3_1, kernel=(3, 3), pad=(1, 1), name=('%s_tower_1' % name), suffix='_conv_1')
    tower_3x3 = Conv(tower_3x3, num_3x3_2, kernel=(3, 3), pad=(1, 1), name=('%s_tower_1' % name), suffix='_conv_2')
    pooling = mx.sym.Pooling(data=data, kernel=(3, 3), stride=(1, 1), pad=(1, 1), pool_type=pool, name=('%s_pool_%s_pool' % (pool, name)))
    cproj = Conv(pooling, proj, name=('%s_tower_2' %  name), suffix='_conv')
    concat = mx.sym.Concat(*[tower_1x1, tower_5x5, tower_3x3, cproj], name='ch_concat_%s_chconcat' % name)
    return concat

def Inception7B(data, num_3x3, num_d3x3_red, num_d3x3_1, num_d3x3_2, pool, name):
    tower_3x3 = Conv(data, num_3x3, kernel=(3, 3), pad=(0, 0), stride=(2, 2), name=('%s_conv' % name))
    tower_d3x3 = Conv(data, num_d3x3_red, name=('%s_tower' % name), suffix='_conv')
    tower_d3x3 = Conv(tower_d3x3, num_d3x3_1, kernel=(3, 3), pad=(1, 1), stride=(1, 1), name=('%s_tower' % name), suffix='_conv_1')
    tower_d3x3 = Conv(tower_d3x3, num_d3x3_2, kernel=(3, 3), pad=(0, 0), stride=(2, 2), name=('%s_tower' % name), suffix='_conv_2')
    pooling = mx.symbol.Pooling(data=data, kernel=(3, 3), stride=(2, 2), pad=(0,0), pool_type="max", name=('max_pool_%s_pool' % name))
    concat = mx.sym.Concat(*[tower_3x3, tower_d3x3, pooling], name='ch_concat_%s_chconcat' % name)
    return concat
```

카카오에서 매월 발행하는 AI 리포트 입니다.

kakao

KAKAO AI REPORT

Vol.01

발행일 | 2017년 3월 21일(2nd Edition)
발행처 | (주)카카오
발행인 | 카카오 정책지원팀
총괄 | 정혜승
편집 | 김대원, 김명수, 정수헌
디자인 | 김성현
메일 | kakaoaireport@kakaocorp.com

COVER
카카오 AI 리포트의 표지에선 AI와 관련된 의미 있는 코드들을 매월 소개할 예정입니다.

Vol.01 코드 | 변민우 dylan.byeon@kakaocorp.com
다음앱 꽃검색에서 사용하는 코드 일부를 발췌/수정한 내용입니다. (http://mxnet.io 코드 인용) '다음앱 꽃검색'은 꽃 이미지 빅데이터를 딥러닝으로 학습하여 이용자가 사진을 찍었을 때 어떤 꽃인지 알려주는 서비스입니다.

감수 1 | 신정규 jshin@lablup.com
소속 : 래블업 대표, 한양대학교 ERICA 컴퓨터 공학과 겸임교수 | 경력 : Google Developer Expert 머신러닝 분야 | 연구 분야 : 복잡계 뇌과학, 기계학습 및 행위자 기반 모형

감수 2 | 최예림 yrchoi@kgu.ac.kr
소속 : 경기대학교 산업경영공학과 교수 | 경력 : 네이버 랩스 연구원 | 연구 분야 : 데이터사이언스, 머신러닝/딥러닝, 휴먼모델링

contents

카카오 AI 리포트를 내면서

어느날 갑자기, 모두 인공지능(Artificial Intelligence·AI)을 이야기합니다. AI가 바꿀 미래, 기술 혁신이 가져올 SF소설 같은 최첨단 상상부터 초지능 로봇에 대한 막연한 두려움, 일자리 걱정까지 그 스펙트럼은 계속 넓어집니다. 알파고와 이세돌의 바둑 대결 이후 아직 1년도 지나지 않았지만 AI는 이미 세상을 바꾸고 있습니다. 정부의 미래 정책 과제의 핵심이 됐고, 기업들의 미래 전략 화두가 됐고, 개인의 일상적 대화 주제가 됐습니다. 글로벌한 논의와 담론도 계속 쏟아집니다.

변화의 속도가 너무 빨라서 방향성에 대한 고민이 이어집니다. 사실 어떻게 대응해야 할지 혼란도 적지 않습니다. 컴퓨터공학자 앨런 케이(Alan Kay)는 “미래를 예측하는 최선의 방법은 스스로 만드는 것” (The best way to predict the future is to invent it)이라고 했지만, 어디서 출발해야 할까요. AI는 하늘에서 뚝 떨어진게 아니라 오랜 연구가 있었습니다. 그리고 최근에는 전 세계에서 관련 연구가 이어지고 있으며, 기술적 연구를 넘어 다양한 분야, 다양한 주체 별로 AI를 탐구하고 미래를 준비하고 있습니다.

카카오는 AI 관련 소식과 정보 가운데, 조금 더 들여다 볼 내용을 정리해서 정기 ‘리포트'로 공유하고자 합니다. 첫 리포트는 다음의 세 가지 내용으로 구성되어 있습니다. 첫째, AI란 무엇인지 알아 봅니다. 인간의 지능을 흉내 내기 위한 그간의 노력들, 역사적 연구 결과들과 중요한 이벤트들을 소개하며, AI가 어떻게 발전해 왔는지 알아 봅니다. 최근 많이 언급되는 ‘딥러닝’의 개념도 정리해봤습니다. 둘째, 최근 미국 및 유럽에서 논의되고 있는 AI와 관련된 제도와 규범 이슈들을 정리했습니다. 셋째, 로봇 윤리의 변천사에 대해서 알아 봅니다. SF소설의 거장 아시모프가 제시한 로봇 3원칙 부터, 일본의 AI R&D 가이드라인, 이른바 ‘아실로마 AI 원칙'까지 역사적 변천을 살펴봤습니다. 올해 기대되는 AI 컨퍼런스와 AI 팟캐스트들도 정리해봤습니다.

이 리포트가 AI에 대한 사회적 관심을 더 높이는 동시에 다양한 논의의 재료로 쓰일 수 있기를 소망합니다. 카카오도 AI 기반 서비스와 비즈니스를 준비하지만, 기술 생태계를 키우고 사회를 바꾸는 것은 모두의 몫입니다. 미래는 AI가 아니라 우리가 만들어 갑니다. 현안을 나누고 지혜를 모으는데 조금이라도 거들겠습니다.

2017년 3월 14일
카카오 정책지원팀 드림

인공지능,
긴 겨울을 보내고


꽃을
피우다.

글 | 정수헌 noah.jung@kakaocorp.com

천체물리학자가 되기를 꿈꾸며 물리학을 공부했다. 하지만 금융 분석에 더욱 흥미를 느끼고 증권사에서 퀀트 애널리스트로 일했다. 한때 모바일게임 개발도 하고 다양한 경험을 쌓았다. 지금은 AI 기술을 대중들에게 쉽게 전달하기 위해 고민하고 있다.

"90년대와 2000년대 초 아무도 신경망(Neural Network)을 주제로 논문을 쓰려고 하지 않았어요. 성과가 나지 않는다는 이유였죠. 학생들에게 논문을 쓰게 하려면 손목을 비틀어야 할 지경이었습니다."[*1]

2014년 겨울 캐나다 몬트리올에서 열린 지능정보처리 시스템 학회(Neural Information Processing Systems, NIPS). 요슈아 벤지오(Yoshua Bengio) 몬트리올대 교수가 지난 시절의 고충을 털어놓았다. 인공신경망(Artificial Neural Network) 개념은 이미 1940년대부터 제기된 내용이었지만, 2000년대까지 AI 연구에서 큰 성과를 보여주지 못했다. 오래된 개념이고, 성과도 없다는 이유로 모든 학회나 저널에서 논문 게재를 거절하곤 했다. 1987년 이후 해마다 겨울에 NIPS 학회에 참석했던 전문가들은 이른바 AI 기술의 겨울(AI winter)이었던 오랜 시절을 함께 돌아봤다. 얀 레쿤(Yann LeCun) 뉴욕대 교수, 제프리 힌튼(Geoffrey Hinton) 토론토대 교수, 앤드류 응(Andrew Ng) 스탠퍼드대 교수 등 지금은 'AI 사대천왕'으로 꼽히는 이들 모두 그 시절 춥고 어려운 시기를 거쳤다. 현재 레쿤 교수는 페이스북, 힌튼 교수는 구글, 응 교수는 바이두에서 AI 핵심 기술 개발을 위해 모셔간 상태다. 벤지오 교수는 IBM, 구글 등과 공동 연구를 하고 있다.[*1 *2]

*1 참고 | http://www.thetalkingmachines.com/blog/?offset=1445518380000 *2 논문 | McCulloch and Walter (1943): A logical calculus of the ideas immanent in nervous activity

인공 지능이란

2016년 알파고와 이세돌의 바둑 대결 이후 AI 붐이 일고 있지만, 전세계적으로 이미 중요한 기술 연구 분야가 된지 오래다 . AI에 대해서 막연하게 SF 영화에 등장하는 가상현실을 상상하지만, 우리는 이미 실생활의 많은 부분에서 AI 기술을 접하고 있다. 로봇청소기, 스마트폰으로 촬영한 사진의 스타일을 변경해주는 어플리케이션, 온라인 게임 플레이에 등장하는 NPC(Non Player Character) 등이 그것이다. SF영화와 소설속에서 등장하는 AI란 과연 무엇을 의미하며, 어떻게 우리의 생활 속에서 이용되고 있는 기술로 발전 하였을까?

지능이란

인간 뿐 아니라 동물을 포함해, 생명체들은 다양한 의사 결정을 필요로 한다. 예컨대 굶주린 반려견에게 두 명이 각각 먹을 것을 내밀며 오라고 손짓하는 상황을 상상해보자. 반려견은 누가 주는 먹이를 먹을지 판단하고 행동해야 한다. 반려견은 현재의 상황, 즉 시각과 후각을 자극하는 먹을 것과 자신이 굶주린 상황에 대해서 인식한다. 그리고, 두 사람이 자신에게 안전하게 먹이만 줄 것인지, 해를 끼칠 지, 과거의 경험, 혹은 다른 요인을 통해서 판단하고 행동할 것이다. 지능은 이처럼 주어진 환경에서 추론하고 결정하고 행동하는 것을 의미한다.

앞의 반려견의 사례 처럼 지적 활동은 다양한 요소들을 포함하고 있다. 코넬 대학의 로버트 스턴버그(Robert Sternberg)[*3] 교수는 지능이 다음과 같이 세 가지 요소로 구성되어 있다는 분류 모델을 제시했다. 첫째, 분석적(Analytic)으로 문제를 해결하기 위해 필요한 사항들을 분석하는 능력. 둘째, 통합적(Synthentic)으로 새로운 정보와 과거의 정보를 통합적으로 고려하여 창의적(Creative) 결론에 도달하는 능력. 셋째, 실용적(Practical)으로 결론에 따른 실제 행동을 취하는 것.

[지능활동]
상황 파악- 입력된 신로를 바탕으로 현 상태를 파악
추론 및 판단- 상황을 바탕으로 어떤 결론에 도달
반응 혹은 행동- 도달한 결론에 맞게 행동

인간 지능을 흉내 내려는 인공지능

AI 연구는 “인간의 지적 활동을 기계도 할 수 있을까?”라는 의문에서 출발했다. 영국의 천재 수학자인 앨런 튜링(Alan Turing)은 0과 1을 이용해서 복잡한 계산을 수행할 수 있다는 것을 입증했다. 이것이 바로 계산기계(Computer)의 시초가 되었다. 튜링이 1950년에 발표한 논문 “Computing Machinery and Intelligence”[*4]에서 튜링의 일대기를 그린 영화 제목과 같은 “The Imitation Game”을 주제로 기계가 인간의 지능을 흉내(Imitation) 낼수 있는 조건을 제시했다. 튜링이 고안한 방법은 다음과 같다. 판별자인 인간 C가 기계 A가 하는 답변과 인간인 B가 하는 답변을 구별하지 못한다면, 기계 A는 인간의 지능을 가지고 있는 것이 아닐까라고 결론 낼 수 있다는 것이다. 이것을 튜링테스트(Turing Test)라고 부르고, 기계의 지능 여부를 판단하는 방식으로 여전히 사용되고 있다.

이와 같이 지능적인 기계를 만드는 엔지니어링 및 과학으로서 AI가 학문 연구의 한 분야로 자리잡게 된 것은 1956년 다트머스 AI 컨퍼런스(Dartmouth Artificial Intelligence Conference)[*5]에서 존 매카시(John McCarthy)가 AI를 ‘기계를 인간 행동의 지식에서와 같이 행동하게 만드는 것’으로 최초로 정의하면서 부터이다.

***3** 논문 | Sternberg, R. J. (1985): Beyond IQ: A triarchic theory of human intelligence ***4** 논문 | A. M. Turing (1950) Computing Machinery and Intelligence. Mind 49: 433-460 ***5** 참고 | https://www.dartmouth.edu/~ai50/homepage.html

인공 지능의 역사

AI가 처음 등장한 다트머스 컨퍼런스는 복잡계이론, 언어처리, 신경망, 입력신호처리, 창의적사고 분야의 다양한 전문가들이 참여한 학술대회였다. 당시 카네기멜론 대학 소속이던 앨런 뉴웰(Allen Newell)과 허버트 A. 사이먼(Herbert A. Simon)[*6] 은 세계 최초의 AI 프로그램인 논리연산가(Logic Theorist)를 선보였다. 논리연산가는 최초로 기계가 논리적 추론(Logical Reasoning)을 할 수 있다는 것을 보여주었다. 당시, 논리연산가는 1903년 알프레드 노스 화이트헤드(Alfred North Whitehead)와 버트런드 러셀(Bertrand Russell)이 쓴 '수학원리'(Principia Mathematica) 두 번째 장에 등장한 52개의 정리(Theorem)중 38개를 증명해 보였다. 논리연산가는 이후 AI 연구에 영향을 주는 중요한 개념들을 제시해 주었다.[*7]

세계 최초의 AI 프로그램, 논리연산가(Logic Theorist)

첫째, 추론을 통한 탐색. | 논리연산가는 공리(Axiom)로 부터 수학적 정리를 이끌어 내기 위해서, 공리로 부터 출발해 발생가능한 대체, 교환, 분리의 경우들을 모두 찾아 가는 방법을 사용하였다.

둘째, 경험적 방법론(Heuristics) 도입. | 발생 가능한 추론들이 다양할 경우 나무의 가지가 너무 많아질 수 있어서 모든 경우를 검색하는 것은 비효율적임을 깨달았다. 여기서 경험적 방법론을 도입하여 효율성을 높여 주었다. 머신러닝에서 특징 설계(Feature Engineering)가 이런 역할을 해준다.

셋째, 정보보처리언어(IPL, Information Processing Language) 개발. | 프로그램 처리를 상징적으로 표현한 언어인 IPL은 이후 존 매카시(John McCarthy)가 만든 리습(Lisp) 프로그래밍 언어의 기초가 되었다. 스탠포드와 다트머스, MIT 교수를 역임한 매카시 교수는 AI란 용어를 1955년 '지능이 있는 기계를 만들기 위한 과학과 공학'이라는 논문에서 처음 기술했다. 리습(Lisp)은 AI 연구에서 논리적 흐름을 상징적으로 표현하는 중요한 프로그래밍 언어로 활용되었다.

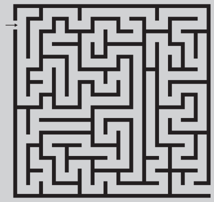
탐색과 추론의 시대

논리연산가(Logic Theorist)에서 제안된 탐색과 추론 방법은 초기 AI 연구에 중요한 개념들을 제시했다. 해결할 할 문제에 대해서 기계가 수행할 일들을 상황별로 대응하는 방식, 이게 바로 탐색(대응)과 추론(상황파악)의 알고리즘이다. 이 알고리즘은 흔히 'If then'(가령 ~ 하면, ~ 한다)이라고도 하는데, 예를 들면 다음과 같다.

[If then rule]
If 어떤 것이 날아 다닌다. then 어떤 것은 새다.

즉, 'If-then rule'은 컴퓨터가 논리적으로 정보를 처리할 수 있도록 원칙(Rule)을 제공해줘서 답을 찾아 가는 과정이다. 미로를 찾는 알고리즘이나, 체커게임(Checkers)을 하는 알고리즘을 개발하는 것 등이 'If-then rule'을 통해 해결할 수 있는 대표적인 문제들이다.

[알고리즘에 대한 이해 - 미로찾기]



그림과 같이 왼쪽 상단의 입구에서 오른쪽 하단의 입구로 찾아가는 미로가 주어졌을 때 길을 어떻게 찾을 수 있을까? 아래와 같은 룰을 설정해서 미로를 찾을 수 있다.

Step 1 (현재 위치 저장)
if 현재의 위치가 갈림길(길이 2가지 이상인 경우) 이면 then 위치를 기억한다.

Step 2 (이동)
지나온 길을 제외하고 길이 남아있는 않은 경우까지 아래를 반복한다.
if 이동할 길이 있으면 then 길이 있는 방향으로 이동한다. 이동한 길은 기억한다.
if 지나 온 길만 남은 경우 Step 3로 넘어간다.

Step 3 (갈림길로 되돌아 가기)
Step 1에서 기억한 위치로 돌아가서 이전에 가지 않은 길로 되돌아 간다.

알고리즘으로 문제를 해결하는 시대의 대표적인 두 가지 방법은 '깊이우선탐색'(Depth First Search)과 '너비우선탐색'(Breadth First Search)이 있다. 깊이우선탐색이란 현재 주어진 문제에서 발생가능한 상황들 중 한 가지를 선택해 깊이있게 해결방법까지 분석해가는 방법이다. 위에서 예제로 제시한 미로찾기 알고리즘은 깊이우선탐색의 예이다. 이에 비해 너비우선탐색은 문제에서 발생가능한 모든 상황을 고려하며 가능성을 넓혀 나가는 것이다. 이렇게 알고리즘을 통한 AI연구의 성과는 인간 체스게임 챔피언과 컴퓨터 체스프로그램과 대결로 이어지게 된다. 1997년 IBM에서 만든 '딥 블루'(Deep Blue)는 당시 세계챔피언인 러시아의 카스파로프(Kasparov)와의 체스 경기에서 승리하게 된다.[*8]

이처럼 초기 AI 기술 연구는 컴퓨터가 주어진 문제를 해결할 수 있도록 인간이 직접 룰을 지정해주는 방식으로 진행됐다. 문제 해결 방법을 논리적으로 설계하고 컴퓨터가 수행할 수 있도록 프로그래밍 하면 지능을 발전시킬 수 있을 것으로 기대했다. 하지만 다양한 데이터를 종합해서 추론하는 좀더 복잡한 문제는 '논리 설계의 어려움' 으로 시도조차 하지 못하게 되었다. 이런 한계속에 1970년 이후 AI 연구 열기가 사그라들면서, 일명 AI의 겨울(AI winter)이 시작됐다.[*9 *10]

시간이 흘러 서서히 AI연구에 반전의 기회가 찾아오고 있었다. 1990년대 인터넷이 보급되기 시작하면서 웹상에 방대한 정보가 쏟아졌다.[*11] 또한, 이렇게 생성되는 정보를 대용량으로 보관할 수 있는 하드디스크(HDD) 의 가격이 급격하게 낮아졌다.[*12] 이른바 '빅 데이터' 시대가 시작된 것이다.

***6** 참고 | 사이먼은 78년 노벨경제학상 수상가이고 뉴웰은 수학자로서 사이먼의 제자 ***7** 참고 | http://history-computer.com/ModernComputer/Software/LogicTheorist.html ***8** 참고 | http://stanford.edu/~cpiech/cs221/apps/deepBlue.html ***9** 책 | Russell, Norvig. Artificial Intelligence a Modern Approach 3rd Edition. Prentice Hall, 2010. ***10** 책 | 마쓰오 유타카. 인공지능과 딥러닝. 동아 엠앤비. 2016. ***11** 책 | Abbate, Janet. Inventing the Internet, Cambridge: MIT Press, 1999. ***12** 참고 | Harddisk Cost / MB - HDD 저장공간당 가격 http://www.mkomo.com/cost-per-gigabyte

머신러닝(Machine Learning), 기계가 스스로 학습 한다.

AI 연구의 1차 붐 시대는 'If-then rule'을 활용한 탐색과 추론을 통한 지능 향상이 목적이었다. 디지털화된 지식과 정보가 빠르게 늘어나는 시대에는 컴퓨터에 지식을 반영하면, 지능을 향상시킬 수 있지 않을까라는 아이디어가 생겨났다. 이렇게 등장한 AI가 전문가 시스템(Expert System)이다. 가령, 의학 분야의 전문가인 의사들은 의학 관련 전문지식을 통해 환자를 진단하고 치료하는 결정을 하게 된다. 전문가 시스템도 이러한 접근 방법에서 고안되었다. 전문가 시스템은 지식 기반(Knowledge Base)에 지식과 정보를 저장하고, 추론 엔진(Inference Engine)에서 기존의 지식을 통해 새로운 지식을 추론하는 두 가지 시스템으로 구성된다. 전문가 시스템은 1970년대 처음 고안되어서 1980년대 크게 유행하게 된다. 궁금한 질문에 대해서 답을 할 수 있는 컴퓨터 시스템을 만들 수 있을까? 이 질문에 답하고자 전문가 시스템을 기반으로 개발된 IBM 왓슨(Watson)은 2011년 미국의 퀴즈쇼 제퍼디(Jeopardy)에 출연해 인간 챔피언들을 이기고 우승을 차지한다.

단순히 책을 보는 것과, 책 속의 내용을 이해하는 것은 다르다. 막 한글을 깨우친 아이에게 칸트(Immanuel Kant)의 '순수이성비판' 책을 읽도록 하면 어떻게 될까? 즉, 컴퓨터에게 지식을 저장하는 것과 컴퓨터가 그 지식을 이해하는 것은 완전히 다른 영역이다. 이를 해결하기 위해 등장한 것이 바로 '머신러닝'이다.

보통 인간은 어떤 문제를 해결한 경험을 토대로 그 다음에 등장하는 문제를 이전보다 개선된 방법으로 해결하려고 노력하게 된다. 머신러닝은 경험적으로 문제를 해결하는 방법을 컴퓨터에 적용한 것이다. 컴퓨터에게 특정 과제(T, Task)를 해결 하면서 그 성과를 측정(P, Performance measure)하는 경험(E, Experience)을 하게 한다. 그야말로 기계를 학습시켜 과제(T) 수행에 대한 측정(P)이 개선되도록 지속적 경험(E)을 수행하는 구조다.[*13]

예들 들면, 야구 경기 승리 전략을 학습하는 머신러닝 프로그램을 가정해 보자. 여기서 과제(T)는 야구 경기를 승리 하는 것. 성과 측정(P)은 득점을 많이 하고 실점을 최소화 하는 것 즉, 득-실 차이를 최대로 하는 것이 될 것이다. 여기서의 경험(E)은 실제 야구 경기를 수행하는 것이다. 프로그램은 경기에서 발생하는 다양한 상황인 투수의 방어율, 그 날의 컨디션, 수비 능력, 타자의 타율, 주루 능력 등 수 많은 조건에 따라서 경험(E)을 훈련하게 된다.

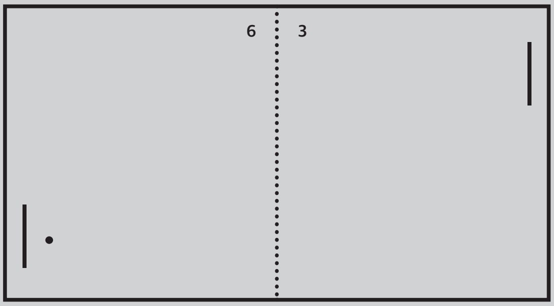
머신러닝에서 사용하는 세 가지 학습방법[*14]

지도학습 (Supervised Learning) | 지도학습은 입력된 데이터에 대한 판단 결과가 명확히 주어진 경우 사용하는 방법이다. 아이들이 사물을 하나씩 익혀 나가는 과정을 보면, 사물을 실제 접하기 전에 그림책을 먼저 보는 경우가 많다. 한 아이가 자동차 종류에 대한 그림책을 보면서 승용차, 버스, 트럭에 대해서 알게 되었다. 이제 밖에 나가서 지나가는 차들을 보면서, "이건 승용차, 저건 버스" 라고 배운 지식을 활용하게 된다. 갑자기 도로에 견인차가 지나가게 되면, 아이는 이전에 배운 트럭과는 다른 모습을 보면서 '저건 트럭인가요?' 하고 묻게 될 것이다.

비지도학습 (Unsupervised Learning) | 비지도학습은 입력된 데이터에 대한 판단 결과가 명확하게 주어지지 않은 경우 사용한다. 답이 정해져 있지 않기 때문에 하나의 결과를 도출할 수 없기에 주로 군집분류(Clustering)에 사용하는 방법이다. 가령, 서울로 출퇴근 하는 사람들의 이동 경로 데이터만을 알고 있는 경우, 기계가 스스로 학습하면서 이들의 사는 지역, 출근 지역을 구분하게 된다. 군집분류 예제 중 하나는 아이리스(Iris data set) 꽃 분류 문제다.[*15] 아이리스 꽃 분류 문제는 150개의 꽃 이미지를 꽃받침(Sepal)의 길이와 폭, 꽃잎(Petal)의 길이와 폭 등 네 가지 특징을 이용해서 총 세 종류의 꽃(Iris Setosa, Iris Versicolour, Iris Virginica)으로 군집분류하는 것이다. (실제 데이터에는 꽃 종류 이름이 주어져서 지도학습으로도 사용되지만, 특징만 가지고 꽃의 종류를 군집분류 하는 비지도학습으로 많이 사용된다.) 지도학습은 입력된 자료 A에 대해서 A이다 라는 답(Label)을 주고 A가 A임을 알 수 있도록 스스로 학습 하는 것이다. 이에 비해 비지도학습은 A와 B의 두가지 입력된 자료가 있는데 이 둘이 A인지 B인지 모르고 단지 둘의 차이를 스스로 학습해서 '다르다'라고 군집분류하는 것이다.

강화 학습 (Reinforcement Learning) | 강화학습이란 주어진 문제의 답이 명확히 떨어지지는 않지만, 결과에 따라서 보상(Reward)과 손실(Penalty)이 주어진다면 이를 통해 보상을 최대화 하는 방향으로 진행하도록 모델을 학습하는 방식이다. 입력된 데이터에 대한 답이 명확하게 주어지지 않는다는 점에서 비지도학습과 유사하지만, 결과를 통해 피드백을 받아 이를 학습에 반영한다는 것이 기존의 학습방식과 다르다. 강화학습은 주로 게임을 플레이 하거나 전략적인 판단을 통해 방향을 설정할 때 활용되는 방식이다. 강화학습으로 해결할 수 있는 쉬운 예제를 들면 '퐁' 게임을 학습하는 경우를 생각할 수 있다. 퐁 게임은 양쪽의 바(Bar)를 상하로 움직이면서 공을 받아 내는 게임이다. 공을 받아치지 못하면 상대방이 득점을 하게 된다. 머신러닝으로 퐁 게임을 학습 하는 경우 플레이 결과에 따라서 받게되는 보상(Reward)이 주어지게 된다. 기계학습은 이렇게 주어진 보상을 프로그램이 최대로 달성할 수 있도록 이루어지게 된다.[*16]

[The Pong Game]



특징 설계가 필요한 머신러닝

빅데이터 시대가 도래하고 SVM(Support Vector Machine), RBM(Restricted Boltzmann Machine)같은 방법론이 발전하면서 머신러닝은 AI연구 분야를 주도하게 된다. 하지만, 머신러닝의 발전에도 한계가 있었는데 바로 '특징 설계'(Feature Engineering)의 문제다.

가령 머신러닝을 이용해 사람 얼굴 이미지 인식을 하는 경우 검은색은 머리고, 눈은 동그랗고, 얼굴의 윤곽은 이미지의 밝기 차이로 구분하는 등 각각의 특징을 통해 다른 사물과 사람 얼굴을 분류하게 된다. 일반적으로 볼 수 있는 사람들의 얼굴로 학습을 진행한 기계에 녹색으로 페이스 페인팅(Face Painting)을 한 사진을 입력한다면 우리는 금방 사람의 얼굴이라고 생각하지만, 기계는 사전에 분류된 얼굴색의 특징과 다르기 때문에 분류에 실패하게 된다.

머신러닝 분류 성능은 사전에 정의된 특징(Feature)과 특징의 중요성을 나타내는 가중치(Weight)를 어떻게 주는냐에 따라 좌우된다. 이렇게 특징을 사전에 정의하는 것은 학습하고자 하는 영역(Domain) 마다 다르게 나타난다. 머신러닝 성능을 높이기 위해서 이미지, 영상, 음성, 텍스트 등 각각의 도메인 별로 특징 설계는 어려운 과제 중 하나였다. 리처드 코엘료(Richert Coelho)가 기술한 책 'Building Machine Learning Systems with Python'에서 '특징 설계'에 관한 챕터를 보면, "잘 선택된 특징으로 만든 간단한 알고리즘이 그다지 잘 선택되지 못한 특징의 최신 알고리즘 보다 좋은 결과가 나온다"고 하고 있다. 결국 머신러닝 성능은 얼마나 효과적인 특징을 찾아내느냐에 따라 달려있다고 할 수 있다. 따라서 머신러닝으로 학습을 할 경우 데이터의 영역(Domain)에 따라서 데이터의 특징 설계를 할 수 있는 전문가들이 필요하다.

***13** 책 | Mitchell, T. (1997). Machine Learning. McGraw Hill. ***14** 참고 | 앤드류 응 교수, 스탠퍼드 강연 https://forum.stanford.edu/events/2011/2011slides/plenary/2011plenaryNg.pdf ***15** 참고 | http://www.cs.upc.edu/~bejar/apren/docum/trans/09-clusterej-eng.pdf ***16** 참고 | http://karpathy.github.io/2016/05/31/rl/

딥러닝(Deep Learning), 인간 뇌의 정보처리 방식을 흉내낸다

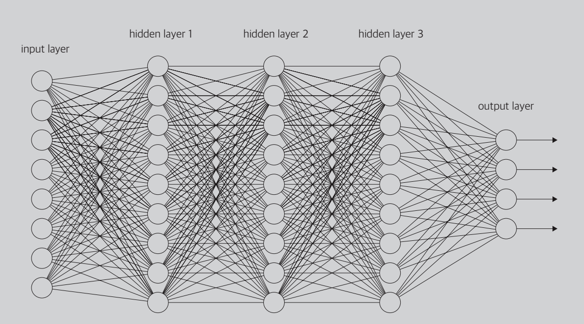
2012년 '이미지넷'(ImageNet)의 이미지 인식 경연대회에서 AI 연구 분야의 흐름을 바꾸는 일이 벌어졌다. 이미지넷은 인터넷의 각 이미지에 이름표를 달아주는 크라우드 소싱프로젝트다. 2010년 부터 'Imagenet Large Scale Visual Recognition Challenge'(ILSVCR)[17] 라는 경연대회를 통해 전세계의 연구 그룹들이 각자의 이미지 인식 기술을 겨루고 있다. 대회 중의 한 분야는 이미지 인식을 통해 1000개의 사물을 구분하는 것이다. 참여 팀들은 알고리즘 기술을 통해 하나의 이미지에 다섯 개의 사물 명칭(Label)을 제시하고 이를 평가하여 정확도를 측정한다.

2012년도 대회에서 토론토대학교의 제프리 힌튼 교수팀의 'SuperVision'은 'Large, Deep Convolutional Neural Network'이라는 알고리즘을 이용해 이미지를 인식했다. 딥러닝 알고리즘의 일종인 CNN을 사용한 힌튼 교수팀은 기존의 방법론을 뛰어넘는 혁신적인 성과를 보여주었다. 2011년도의 1위팀과 2위팀의 오류률 차이는 1위(25.7%)와 2위(31.0%) 에서 2012년도 'SuperVision'은 16.4% 로 2위팀의 26.1%와의 큰 차이로 우승하게 된다.[18][19]

인간의 뇌는 뛰어난 정보처리 시스템이다. 시각, 후각, 촉각, 청각, 미각 등 다양한 감각 기관으로 입력된 정보를 종합적으로 판단해 즉각적인 반응을 내놓게 된다. 수 많은 뉴런들의 연결로 구성되어 있는 뇌의 구조를 '신경망'(Neural Network)이라고 한다. 하나의 뉴런은 다양한 뉴런들과 연결되어있고, 시냅스(Synapse)를 통해 신호를 주고 받으며 정보를 전달하게 된다. 뉴런에 임계치(Threshold) 이상의 신호가 입력되면, 그 뉴런은 활성화(Activation)되면서 시냅스를 통해 신호를 전달하고 임계치를 넘지 않은 경우 뉴런이 활성화 되지 않아 신호가 전달되지 못한다.

딥러닝은 뇌의 정보 전달 방식과 유사하게 신경망 구조를 여러 층으로 깊이(Deep)있게 구성하여 학습을 진행하는 것이다.

[Deep Neural Network 구조][20]



가령 여러 사진에서 고양이 이미지를 인식하는 학습을 진행한다고 가정해보자.[21] 기존 머신러닝 방법론에서는 특징 설계를 통해서 다른 동물과 고양이의 얼굴을 잘 구분할 수 있는 특징을 사전에 추출(Feature Extraction)하고 추출된 특징을 바탕으로 분류하며 학습을 진행한다. 하지만, 딥러닝은 정보를 구분할 수 있는 최소한의 단위, 이미지의 경우 픽셀 데이터를 입력하면, 입력된 값과 출력된 결과물의 오차가 최소화 되는 방향으로 네트워크가 스스로 학습한다. 중간에 특징 추출 과정이 사라지고 입력과 출력의 데이터만 주어지면 학습을 할 수 있기 때문에 'End to End'(E2E) 학습이라고도 부른다.

'인공신경망'(Artificial Neural Network)의 역사는 1940년대까지 거슬러 올라 간다. 프랭크 로젠블라트(Frank Rosenblatt) 코넬대 교수는 입력신호를 연산해서 출력하는 퍼셉트론(Perceptron)이라는 개념을 1957년에 제시했다. 퍼셉트론은 비선형연산에 대해선 작동하지 않는 한계로 금방 잊혀졌지만, 1980년대 퍼셉트론을 다층구조(Multi-layer)로 연결하면, 비선형연산도 가능하다는 사실이 발견됐다. 그러나 다시 'AI 겨울' 동안 관련 연구에서 완전히 배제됐다. 당시에는 어쩔 수 없는 시대적 한계가 있었다.

첫째, 층을 깊이(Deep) 있게 쌓아야 AI의 성능이 개선되지만, 계산할 양이 많아 지게 되고 이는 당시의 컴퓨팅 능력으로 극복하기 힘든 한계였다.

둘째, 층이 깊어 진다는 것은 층간의 연결 변수들이 증가한다는 것을 의미하는데 이는 학습을 통해 찾아야 할 미지수가 늘어나는 것을 의미한다. 우리는 중학교때, 변수가 두 개 x, y 인 연립 방정식에 대해서 배운다. 즉 두 개의 변수를 찾기 위해선 두 개의 방정식이 필요한 것처럼 찾아야할 변수가 늘어날수록 입력해야할 데이타가 많이 필요하다.

시간이 약이 되었을까? 2000년대에 이 두가지 문제가 해결되기 시작한다. 빅데이터 시대를 맞이하여 데이터는 넘쳐나기 시작했고, 컴퓨터 H/W의 성능 개선 뿐 아니라 병렬연산, GPU연산 등의 방법이 등장하면서 연산속도가 극적으로 개선되었다. 2012년 힌튼 교수팀 'SuperVision'이 사용한 모델은 65만개의 뉴런과 6000만 개의 변수, 6억3000만 개의 네트워크 연결로 구성되었다.

딥러닝(Deep Learning) 'Deep'의 의미는?

딥러닝(Deep Learning)의 'Deep'은 다양한 측면의 의미를 가지고 있다. 대부분의 AI연구자들에게는 인공신명망의 은닉층(Hidden Layer)이 2개층 이상인 경우 Deep(깊은)이라고 생각한다. 이러한 정의는 2개층 이상으로 신경망을 구성할 때, 성공적인 학습과 예측을 수행한다는 것을 의미한다. 실제 2012년 'SuperVision'팀이 사용한 신경망은 7개의 은닉층을 가지고 있었다. 최근 이미지 인식

연구에서 마이크로소프트 연구팀 'ResNet' 모델은 무려 150개의 은닉층을 사용해서 예측 정확도를 높였다.(오류율 3.57%)*22 또 다른 의미로 딥러닝(Deep Learning)은 '심화학습'을 뜻하기도 한다. 인공신경망 학습에서는 인간이 수행하는 특징설계가 필요치 않기 때문에 기계가 스스로 입력된 데이터의 특징을 찾기 위한 학습 과정을 거친다. 이러한 과정을 심화학습이라고 표현할 수 있다. 인공신경망 연구가 AI의 겨울을 끝내고 새로운 도약의 시기를 맞이할 수 있었던 이유는 데이터, 연결, 하드웨어 개선 등 여러가지 변수가 함께 새로운 시대를 열었기 때문이다.

***17** 참고 | http://cs.stanford.edu/people/karpathy/cnnembed/cnn_embed_1k.jpg ***18** 참고 | 2012년 이미지넷 결과, http://image-net.org/challenges/LSVRC/2012/analysis ***19** 논문 | ImageNet Classification with Deep Convolutional Neural Networks - Krizhevsky et al. NIPS. 2012. ***20** 참고 | http://stats.stackexchange.com/questions/234891/difference-between-convolution-neural-network-and-deep-learning ***21** 참고 | https://static.googleusercontent.com/media/research.google.com/en//archive/unsupervised_icml2012.pdf ***22** 참고 | http://kaiminghe.com/icml16tutorial/icml2016_tutorial_deep_residual_networks_kaiminghe.pdf

인공지능 기술의 현재

우리는 2016년 이세돌을 이긴 알파고를 AI의 대명사로 생각하고 있다. 딥러닝 방법론이 AI 연구에서 뛰어난 성과를 보이기 시작한 이후 다양한 분야에서 인간지능에 도달했거나 혹은 인간을 뛰어넘는 AI들이 출연하고 있다.

인간을 넘어서는 AI시대가 오고 있다.
최근 카네기멜론 대학 연구팀에서 만든 AI 프로그램은 '텍사스 홀덤 포커 게임'에서 세계챔피언을 이겼다.[*23] 텍사스 홀덤 게임은 미국에서 가장 많이 즐기는 포커 게임으로 매년 수십억원의 상금이 걸린 세계 대회가 열리고 있다. 포커는 운이 승패를 좌우하는 경향이 크다. 하지만 텍사스 홀덤 포커 대회는 여러번의 경기를 통해 가장 많은 게임 칩을 획득한 참가자가 우승하는 방식이다. 즉, 한번의 게임 승패 보다 전체적인 게임 운영 전략이 중요한 요소로 작용한다. 이기는 게임의 베팅을 높이고 지는 게임의 베팅을 낮추는 식의 전략적 판단이 필요한 것이다. 이러한 게임에서 AI가 세계 챔피언을 이겼다는 것은 전략적 의사 결정까지도 일정 수준을 뛰어 넘었다는 의미다.

페이스북에 사진을 올려보면, 얼굴을 자동으로 인식, 이 사람이 맞는지 물어보며 자동 태그하는 기능을 경험할 수 있다. 페이스북이 2014년에 선보인 딥페이스(DeepFace)라는 기술로 사진에서 사람을 인식해 누구인지 알려주는 서비스다. 2015년 구글은 페이스넷(FaceNet)이라는 얼굴 인식 시스템을 발표 했는데 연구 결과 99.96%의 인식률을 보여준다고 한다. 인간이 사진을 보고 사람을 구분하는 정확도가 평균적으로 97.53%임을 감안하면 인간보다 훨씬 높은 수준을 달성한 것이다. 특히, 인간의 경우 사진이 흐리거나, 조명이 너무 밝거나 어둡거나, 사진의 각도가 달라지게 되면서 얼굴의 일부분만 보이는 경우 오류를 범하기 쉽지만, AI는 대부분 높은 정확도로 얼굴을 인식하고 있다.[*24 *25]

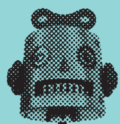
당뇨망막병증(Diabetic Retinopathy, DR)은 당뇨병의 합병증으로 발생하는 병으로 망막의 미세혈관 손상으로 실명에 이르게 되는 병이다. 전세계적으로 발생하는 실명의 원인 중 가장 높은 비중을 차지하고 있는 질병이다. 초기에 진단된 DR은 치료를 통해서 악화를 막을 수 있다. 기존에 고도로 훈련된 안과전문의들이 망막을 스캔한 사진을 직접 관찰하면서 진단했다. 구글 연구팀은 2016년 11월 AI가 97~98% 수준으로 DR을 진단할 수 있다는 논문을 발표했다. 구글은 AI를 훈련시키기 위해서 수년간 미국 전역의 병원에서 약 12만개의 안구 이미지를 분석했다. 이 기술은 DR의 초기 진단을 좀더 쉽고 편하게 이용할 수 있게 해주면서 많은 환자들을 치료할 수 있을 것으로 기대 되고 있다.[*26]

우리가 인지하지 못한 사이 이미 많은 분야에서 AI는 인간에 가깝거나 인간을 넘어서는 능력을 보이고 있다. 1997년도에 AI와 경쟁하여 체스를 이긴 마지막 인간이 러시아의 카스파로프(Kasparov)였다면, 바둑을 이긴 마지막 인간은 이세돌이 될 것이다. 이미 다양한 분야에서 인간의 감각기관중 시각, 청각을 대체하기 위해 AI가 사용되어 인간 수준의 지각 능력을 보여주고 있다. 또한, 복잡한 계산과 전략적 추론에서조차도 인간의 능력을 넘어서고 있다.

앞으로 우리는 '카카오 AI 리포트' 를 통해 다양한 분야의 AI 사례들을 소개할 예정이다.

***23** 참고 | http://www.digitaltrends.com/computing/texas-holdem-libratus-ai-defeats-humans
***24** 참고 | http://www.dailymail.co.uk/sciencetech/article-3003053/Google-claims-FaceNet-perfected-recognising-human-faces-accurate-99-96-time.html ***25** 참고 | 2016년 기준 이미지 인식률, Facebook DeepFace (97.35%) 인간 (97.53%) Google (99.96%) ***26** 참고 | https://research.google.com/teams/brain/healthcare

인간과 로봇의 공존시대 열리나.



글 | 김명수 allen.kim@kakaocorp.com

우완 정통파 투수를 좋아한다. 기자로 일하며 직구로 승부하는 이타적 삶을 꿈꾸다 미국법을 공부한 뒤 변화구 섞어 던지는 법을 익혔다. 기업법무와 해외투자 분야를 거쳐 정책 업무를 맡고 있다. 좋은 규제와 정책은 세상을 바꾸고 사람들을 더 행복하게 해줄 수 있다는 믿음을 갖고 있다.

2015년 2월 네덜란드에서는 트위터 '봇(Bot) 계정'이 살인을 예고하는 듯한 문장을 트윗해 경찰이 조사에 나섰다. 이 봇 계정은 계정 주인이 과거에 작성한 트윗들에서 무작위로 뽑아낸 단어들로 문장을 자동 생성한 뒤 다른 봇 계정과 공유하도록 설계됐다. 경찰은 문제를 일으킨 봇 계정 소유자에게 책임이 있다는 경고를 한 뒤 봇 삭제를 요청하는 것으로 사건을 마무리 지었다. 이 사건은 결국 봇을 '처벌(삭제)'하는 것으로 끝났지만 '봇이 알고리듬에 따라 자동 생성한 살인 예고에 대해 누가 법적 책임을 질 것인가'란 질문은 여전히 남는다. 가까운 미래에 인공지능(Artificial Intelligence·AI) 기술이 더욱 발전해 로봇이 스스로 학습하고 판단하는 능력을 갖추고 '전자 인간(Electronic Person)'이라는 법적 지위까지 얻게 되면 로봇의 실제 범행 의도(Mens Rea)까지 따져봐야 할지 모른다.

인공지능 어떻게 정의할 것인가

자율적으로 학습하고 실행하는 능력을 갖춘 AI의 출현은 역으로 인간이 기계에 대한 통제권을 잃게 되는 상황을 의미한다. 인간은 알고리듬 설계를 통해 초기 통제권을 갖지만 일단 AI 가 학습 능력을 획득하면 어떤 과정을 거쳐 해결책을 도출하는지 파악하기 힘들 수 있다. 최악의 경우, 인간의 고통을 최소화기 위해 설계된 AI 가 '인간은 그 특성상 천국에서도 고통을 받는다'라는 사실을 학습한 뒤 인류 멸종이 인간의 고통을 없애는 최적의 솔루션이라 판단할 수 있다. 글자 그대로의 의미는 정확히 이해하지만 인간의 주관적인 의도나 그러한 의도를 해석하는 행위에 대해서는 기계가 무관심할 수 있기 때문이다.[*1]

국내에서는 2016년 3월 알파고(AlphaGo)와 이세돌 9단과의 대결이 알파고의 승리로 끝나면서 한때 'AI 공포증(AI Phobia)'이란 단어가 등장할 정도로 AI 에 대한 두려움이 커졌다. AI 전문가들이 '알파고는 아직 AI 단계에 미치지 못한다'며 진화에 나섰지만,[*2] 한편에서는 AI가 기술적 특이점인 '싱귤래리티(Singularity)'를 거쳐 '캄브리아 폭발기'에 비견될만큼 비약적으로 발전할 것이란 예측도 나오고 있다.[*3]

엄청난 잠재력과 파급력을 가진 AI 기술에 대한 불안감과 두려움은 'AI 기술을 어떻게 규제할 것인가'란 질문으로 이어진다. 자율주행차량, 드론, 헬스케어 등 지능정보화 기술을 적용한 응용 서비스 분야에서는 상업화 시점이 가까워지면서 사고시 책임 소재, 알고리듬 설계 규제, 제조물 책임법, 책임 보험, 로봇 창작물의 저작권 등 새로운 법적 이슈들이 활발하게 논의되고 있다. 반면 인간을 닮은 '강한 AI '에 대응하기 위한 논의들은 인류가 그동안 생각해보지 못했던 새로운 윤리, 철학, 존재론적 이슈들까지 포섭하며 더욱 광범위하게 확대되고 있다.

법과 규제의 출발점은 규율 대상을 '정의'하는 데서 시작한다. 'AI 를 어떻게 규제할 것인가' 질문 역시 규율 대상인 AI 를 어떻게 정의할 것인가에서 출발한다. 현재 AI 분야에서는 모든 전문가들이 동의할만한 정의가 존재하지 않는다. 다만 AI 분야의 교과서로 알려진 '인공지능: 현대적 접근'에서 스튜어트 러셀 교수와 피터 노빅은 인공지능을 정의하려는 시도들을 '인간처럼 사고하고, 인간처럼 행동하고, 합리적으로 사고하고, 합리적으로 행동하는' 네 가지 특징별로 분류한 뒤, '합리적으로 행동하는: 합리적인 에이전트(agent) 접근법'을 채택해 인공지능을 분석했다. 반면 미국 조지메이슨대 매튜 슈어러 교수는 규제 관점에서 볼 때 '특정 기능을 수행할 수 있는-' 식의 목적 지향적(Goal-oriented) 정의는 AI 처럼 급격한 기술 변화가 예상되는 분야에서는 적절하지 않다고 설명한다. 따라서 정기적으로 기술 변화 추이를 점검하며 AI 의 정의를 점진적으로 변화시켜가는 것이 규제 목적에 더욱 부합하는 대안이 될 수 있다.

미국과 EU의 로봇 및 인공지능 규제 동향

AI 과 관련한 정책 및 규제 이슈에는 AI 의 개념, 데이터, 프라이버시, 알고리즘 편향, 일자리, 거버넌스, 법적 책임, 로봇 법인격 등 다양한 이슈들이 포함된다. 미국, 유럽연합(EU) 등 기술 선진국들은 AI 및 로봇산업의 진흥을 위해 기존의 법, 제도, 규제 체제를 새롭게 정비하거나, 폭발적인 기술 혁신에 대비한 장기 과제로 새로운 규범 및 윤리 가이드라인의 도입을 준비하고 있다. 미국과 EU는 AI 가 초래할 변화에 대응할 수 있는 법적, 제도적 틀을 마련해야 한다는 필요성에는 의견을 같이 하지만 구체적인 대응 방식에서는 다소 차이를 드러낸다. EU가 AI 를 규율할 법제적, 윤리적 대응에 좀더 적극적으로 나서는 반면 미국은 AI 기술이 가져올 안전, 공정성 문제, 장기 투자를 위한 정부의 역할 등에 우선 치중할 것을 주문하는 등 주로 공익, 공정성, 책임성 확보를 최우선 가치로 제시하고 있다. 또한 구체적인 대응 방식에 있어서도 미국은 윤리, 법제적 대응의 필요성을 인정하면서도 구체적인 법제화보다는 지속적인 모니터링을 우선 권고하고 있다.

***1** 논문 | Stuart J. Russell & Peter Norvig, 'Artificial Intelligence: A Modern Approach' 1037 (3d ed. 2010). ***2** 참고 | Jean-Christophe Baillie 'Why AlphaGo Is Not AI', IEEE Spectrum 2016년 3월 17일 게재 http://spectrum.ieee.org/automaton/robotics/artificial-intelligence/why-alphago-is-not-ai ***3** 논문 | Gill A. Pratt, 'Is a Cambrian Explosion Coming for Robotics?' Journal of Economic Perspectives, vol. 29, No. 3, Summer 2015, pp.51-60.

EU의 로봇법 프로젝트

EU는 2012년부터 유럽 4개국(이탈리아, 네덜란드, 영국, 독일)의 법률, 공학, 철학 전문가들이 참여하는 컨소시엄 형태인 로봇법(Robolaw) 프로젝트를 통해 로봇과 관련한 법적·윤리적 이슈를 연구한 뒤 2014년 9월 결과 보고서 'D6.2 로봇규제 가이드라인(D6.2 Guidelines on Regulating Robotics)'*4을 도출했다. 유럽연합의 로봇법 및 규제 가이드라인은 자율주행자동차, 수술 로봇, 로봇 인공기관, 돌봄 로봇 등 상용화에 근접한 기술들을 사례별(Case-by-case)로 분석함으로써 구체적이고 기능적인 맥락에서 로봇 관련 규범 체계를 정립하기 위한 목적에서 추진됐다. 따라서 새로운 규제의 도입보다는 안전, 법적 책임, 지적재산권, 프라이버시, 데이터, 로봇의 계약 체결 능력 등 주요 이슈별로 기존 법률의 적용 가능성을 우선 검토하거나, 법률, 기술, 윤리, 의료 등 다양한 분야의 전문가들이 참여하는 다자간 학제적인 접근 방식을 취했다. 자율성을 가진 로봇의 예상치 못한 행동으로 초래된 법적 책임에 관한 원칙들에 관해서는 혁신 장려와 위험 규제라는 상반된 가치들을 균형있게 조화시키며 논의할 것을 주문했다. 또한 인공지능 기술이 인간의 기본권 보호라는 대원칙을 훼손시키지 않아야 하며, 인간 역량 강화라는 더 큰 목적을 목적을 달성할 수 있어야 한다고 강조했다. EU는 앞선 2007년 유럽로봇연구네트워크(EURON)의 ICRA(IEEE International Conference on Robotics and Automation) 회의를 통해서는 'EURON 로봇윤리 로드맵'을 발표하고 로봇윤리에 관한 실용적 접근 전략 차원의 권고안을 제시한 바 있다.

EU 회원국들이 지난 수년 간 AI 와 로봇 분야에서 쌓아온 연구 내용들은 2017년 1월 유럽의회 법사위원회가 작성한 결의안(Draft Report with Recommendations to the Commission on Civil Law Rules on Robotics)*5에 반영됐다. 'Robolaw' 프로젝트가 유럽연합 집행위원회 주도로 이뤄졌다면 이번 결의안은 유럽의회 주도로 진행됐다. 결의안 작성을 이끈 법사위원회 부위원장 매디 델보에 따르면*6, 결의안에서 언급된 로봇이란 '센서나 주변 환경과의 상호작용을 통해 얻은 데이터를 분석해 자율성을 획득하는 능력, 자율적인 학습과 적응 능력을 갖춘 물리적인 기계'로 정의되며, 군사용을 제외한 민간용 자율주행차, 드론, 산업용 로봇, 돌봄 로봇, 엔터테인먼트 로봇 등이 이 범주에 포함된다.

유럽의회는 이 결의안에서 일상 생활에서 로봇의 영향력이 점차 확대되는 가운데 로봇이 인간에게 도움을 주는 존재임을 명확히 하기 위해 EU내 강력한 법적 프레임워크가 필요하다고 촉구했다. 특히 로봇에 '전자 인간(Electronic Person)'의 지위를 부여해 로봇으로 인한 피해 발생시 법적 책임 소재를 명확히 하는 법적 기틀을 마련했다. 자율주행자동차 상용화를 앞두고 사고시 피해자가 충분히 보상을 받을 수 있도록 지원하는 책임 보험과 전용 기금의 도입도 포함됐다. 로봇 세금(Robot Tax)에 관해선 명확한 용어로 지칭하지는 않았지만 로봇 이용으로 발생하는 금전적인 혜택을 피해 보상 전용 기금에 할당하는 안이 제시됐다. 결의안에는 기술, 윤리, 규제 분야에서 로봇과 AI에 관한 전문성을 갖춘 전문기관의 설립을 촉구하는 내용도 담겼다. 아울러 로봇 제작자가 비상 상황에서 로봇을 즉시 멈출 수 있는 '킬 스위치(Kill Switch)' 를 설계 시점부터 적용할 것을 제안했다. 결의안에 앞서 2016년 5월 법사위원회가 작성한 보고서(Draft Report)는 로봇 도입 확산에 대처하기 위해 로봇 공학자와 설계자, 생산자, 이용자들이 준수해야 할 윤리 강령(Ethical Code of Conduct), 연구기관의 투명성과 책임성, 설계 라이선스 등 분야별 규범 및 규제 원칙들을 제시했다.

[유럽의회의 로봇 관련 결의안 주요 내용]

구분	주요 내용
로봇 등록제	첨단 로봇의 경우 로봇 분류 체계에 따라 등록제 도입
저작권 및 프라이버시	로봇 창작물의 저작권 보호 기준을 마련하고 정책 입안 시점부터 프라이버시 및 데이터 보호 방안을 적용
법적 책임	- 향후 10~15년 간 로봇 및 인공지능 발전으로 초래되는 법적 이슈를 다루기 위한 입법안을 마련할 것을 EU 집행위원회에 요청. - 로봇 때문에 발생한 피해라는 이유만으로 보상 가능한 손해의 종류나 정도를 제한하거나 보상의 종류를 제한할 수 없음. - 로봇이 초래한 피해에 대해선 엄격한 책임 원칙(strict liability)을 적용하고, 로봇의 행위와 손해간 일반적인 인과관계만 입증하면 보상 요건 충족하도록 설정. - 로봇이 초래한 손해의 보상, 로봇이 자율적으로 결정한 사건 등을 다루기 위해 로봇에게 '전자 인간'의 지위를 부여.
로봇세	자율 로봇 제조사가 책임 보험에 가입하되 보험 미적용 사고의 보상을 위해 기금을 조성. 로봇이 행한 서비스의 대가로 로봇에게 지급되는 금전적 대가를 이 기금에 이전하는 안이 제시됨(*일종의 '로봇세'로 의원총회 표결 과정에서 부결됨).
윤리원칙	로봇 디자인, 개발, 연구, 이용시 적용되는 윤리강령 원칙들을 담은 로봇 헌장을 향후 입법시 고려할 것
노동시장 및 기본소득	로봇공학 및 인공지능 발전이 노동시장에 미칠 영향을 감안해 세제, 사회안전망, 기본소득 등에 대한 고려 필요 (*기본소득 내용은 의원총회 표결 과정에서 부결됨)

유럽의회는 지난 2월 16일 의원총회를 열어 로봇세와 기본소득에 관한 내용은 부결시킨 채 이 결의안을 통과시켰다.

EU 의회는 로봇 도입 확산으로 일자리를 잃는 노동자 재훈련과 사고 보상 목적의 기금 마련을 위한 ‘로봇세’ 도입에는 반대했지만 로봇과 관련한 윤리적, 법적 책임성 문제를 해결하기 위한 법제화 필요성은 인정했다. 이에 따라 EU 집행위원회는 AI와 로봇, 그리고 자율주행자동차 등에 관한 법률적, 윤리적 문제 검토에 착수할 전망이다.

자율주행차 분야의 경우, EU는 2009년부터 2012년까지 볼보, 리카르도 등 유럽 7개 기업이 참여하는 SARTRE(Safe Road Trains For The Environment) 프로젝트를 통해 자율주행 자동차 적용을 위한 각종 규제 및 규칙을 점검했다. 이어 2012년 완전 자율주행버스 공동개발 프로젝트인 시티모빌2(CityMobil2), 2014년 'AdaptIVe' 프로젝트를 통해 자율주행자동차 상용화를 위한 법적 검토를 실시했다. EU는 자율주행 관련 입법화를 적극 추진하고 있지는 않지만 자율주행 자동차가 유엔의 도로교통에 관한 비엔나 협약 등 EU 협약국 간 표준화된 교통 규칙에 위배되는지 여부를 검토해왔다. 비엔나 협약은 1968년 UN에서 협약국 간 교통 법규를 표준화하기 위해 제정된 협약으로 이동중인 차량에는 운전자가 반드시 탑승해 있어야 하고, 모든 차량은 운전자 통제 하에 있어야 한다는 조건 등을 두고 있다.

***4** 참고 | RoboLaw(2014), ‘D6.2 ‘Guidelines on Regulating Robotics’ 2017.2.20 last accessed at http://www.robolaw.eu/RoboLaw_files/documents/robolaw_d6.2_guidelinesregulatingrobotics_20140922.pdf ***5** 참고 | EU의회 법사위원회의 결의안. http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//NONSGML+COMPARL+PE-582.443+01+DOC+PDF+V0//EN&language=EN ***6** 참고 | EU의회 뉴스. European Parliament news, ‘Rise of the robots: Mady Delvaux on why their use should be regulated’, 2017.2.20 last accessed at: http://www.europarl.europa.eu/news/en/news-room/20170109STO57505/rise-of-the-robots-mady-delvaux-on-why-their-use-should-be-regulated

미국 백악관의 인공지능 보고서

미국은 2016년 백악관에 ‘기계학습 및 인공지능 소위원회(Subcommittee on Machine Learning and Artificial Intelligence)’를 신설했다. 이 소위원회는 AI 개발 상황과 성과를 모니터링한 뒤 그 결과를 국가과학기술회의(National Science and Technology Council)에 보고하는 역할을 맡고 있다. 소위원회는 2016년 5월부터 7월 사이 네 차례에 걸쳐 AI 가 일자리, 경제, 안전, 규제 분야에 미치는 영향을 공개적으로 토론하는 일련의 워크숍을 개최했다.

미국의 AI 전문가들은 백악관이 주도한 이 토론회를 통해 AI 는 아직 개발 중이므로 섣부른 정부 개입은 오히려 안전하고 책임있는 기계를 개발하는데 방해 요인으로 작용할 수 있으며, 대신 프로그램이나 알고리듬을 설계하는 사람들에게 더욱 높은 도덕적, 윤리적 책임을 부과해야 한다는 의견을 제시했다. 미국 정부 역시 상용화를 앞둔 자율주행차, 드론, 암 진단 분석 시스템 등에 대해서는 규제와 감독 기능을 집중하는 반면 AI 분야에서는 섣부른 법제화나 규제 도입보다는 기존 법적 원칙들을 탄력적으로 적용하거나 공정성, 책임성, 투명성, 적법절차와 같이 정책적인 방향성을 제시하는 데 더욱 주력하고 있다.

백악관 국가과학기술위원회(NSTC)는 이런 활동을 바탕으로 2016년 10월 ‘인공지능 국가연구 개발 전략 계획’, ‘인공지능의 미래를 위한 준비’ 보고서를 발간한 데 이어 12월 ‘인공지능, 자동화, 그리고 경제’ 보고서를 발간하며 AI 기술의 연구개발 및 정책 방향을 제시했다. 특히 AI 분야 장기 투자를 위한 연구개발 지원, AI로 인해 초래되는 사회적 안전 및 공정성 문제에 우선 대처할 것 등 공익 보호와 공정성, 책임성, 투명성 확보를 최우선 가치로 제시했다. 구체적인 대응 방법보다는 AI 발전을 지속적으로 모니터링하고 기초적이고 장기적인 인공지능 연구에 우선 순위를 설정할 것을 주문하는 등 전반적인 정책 방향성을 제시하는데 주력하는 모양새다.

행정부와 입법부가 로봇 관련 법제화에 적극적인 유럽연합과

달리 미국에서는 산업계, 학계 등 민간 분야를 중심으로 '약한 AI' 기술의 발전과 진흥을 위한 규제 패러다임 확립 요구가 강하게 제기되고 있다. 이들은 당장 '강한 AI '의 위험을 걱정하기보다는 상용화에 근접한 '약한 AI ' 기술들이 사회에 미치는 영향과 법적 공백 상태가 더욱 시급한 문제라는 인식을 갖고 있다. 나아가 빅데이터, 프라이버시, 인간의 기계 의존 문제, 일자리 이슈들에 관해 정책, 법률, 과학 등 다양한 분야의 전문가들이 참여하는 규제 패러다임 형성을 요청하고 있다. 미국 실리콘 밸리에서는 민간 주요 기업들과 인공지능 관련 기업들이 주축이 돼 싱크 탱크(Think Tank)를 설립하고 학계와 연계한 AI 연구를 수행하고 있다. 구글은 딥마인드(DeepMind) 의 요청에 따라 신기술 검토를 위한 윤리위원회를 설치했고, 2014년 설립된 미국 보스톤 소재 비영리 연구단체인 '삶의 미래 연구소(Future of Life Institute, FLI)'는 AI 의 잠재적인 위험성과 편익에 관한 연구를 지원하고 있다.

미국의 법현실주의적 접근

미국 법조계나 학계에는 기존의 사이버법(Cyberlaw)이나 민사법 등 전통적인 법리들을 발전시켜 인공지능 시대에 대처하려는 노력들도 존재한다. 워싱턴대 로스쿨의 라이언 칼로(Lyan Calo) 교수는 로봇기술이 신체화(Embodiment), 창발성(Emergence) 사회적 유의성(Social Valence)을 가진다고 설명하면서, 일명 IT법을 통해 발전되어온 인터넷 또는 컴퓨터 관련 법리들을 인공지능 시대에 더욱 확장해 적용시킬 수 있다고 주장한다.[*7] 칼로(Ryan Calo, 워싱턴대) 교수는 'Robots in American Law'(2016) 논문에서 로봇과 인공지능 개념을 언급한 과거 법원 판결들을 분석한 뒤 로봇을 인간의 소유물 같은 단순 객체가 아닌 인간을 대리하거나 인간의 권리를 확장하는 법적 주체로 여기는 논의들이 과거에도 이뤄져 왔다고 설명한다. 즉, 로봇을 위해 완전히 새로운 법적, 규제적 개념을 탐구하기보다는 기존의 법적 논의들을 다시 세밀하게 분석해볼 필요가 있다는 주장이다.[*8] 한편, 미국 예일대학 로스쿨의 잭 볼킨(Jack Balkin) 교수는 미국 법조계의 법현실주의(Law Realism) 흐름에 입각해 칼로의 설명은 로봇 기술의 본질을 미리 확정한 채 논의를 전개하는 한계를 가진다고 지적한 뒤 기술은 사회적으로 어떻게 활용되는지에 따라 다르게 평가할 수 있다고 설명한다. 볼킨 교수에 따르면, 로봇 및 인공지능과 인간들의 상호 작용은 우리가 자각하지 못할지라도 실제로는 '로봇 개발자들과 우리 (인간들) 간'의 상호작용에 해당한다.[*9]

***7** 논문 | Ryan Calo, "Robotics and the Lessons of Cyberlaw", California Law Review 103, 2015. ***8** 논문 | Ryan Calo, "Robots in American Law", Legal Studies Research Paper No. 2016-04 pp. 42-44 ***9** 논문 | Jack M. balkin, "The Path of Robotics Law", California Law Review Circuit 6, 2015.

왜 다시 로봇윤리인가

AI가 인간 한계를 뛰어넘어 인간을 압도하는 능력을 갖게 되면 사실상 인류를 대체할 수 있다는 실존적인 위협은 AI 윤리와 같은 규범적 논의를 더욱 광범위하게 확장시킨다. 인간과 흡사한 AI 가 출현할 경우 결국 자율성을 갖춘 인공지능을 제어할 수 있는 가장 보편적인 방식은 윤리나 규범이기 때문이다. 프린스턴 대학의 피터 아사로(Peter Asaro) 교수에 따르면 로봇은 파괴적 혁신 기술이기 때문에 기존의 법과 정책으로 규제를 하는 데는 한계가 있으며, 로봇과 AI 를 위한 새로운 법률과 정책을 만들기 위해서는 우선 사회 규범을 먼저 형성해야 한다. 달리 말하자면, 사회 규범 없이 정치적 의지만으로는 법과 정책을 만들 수 없으며, 다양한 분야의 전문가들 논의를 통한 공론화가 선행해야 한다.*10

스스로 판단할 수 있는 기계의 출현은 다양한 윤리적 문제를 발생시킬 가능성이 매우 높다. AI 규범과 관련한 법적, 윤리적, 철학적 쟁점은 크게 i) AI 및 로봇의 설계, 생산, 이용 단계에 적용되는 윤리적 과제, ii) AI 및 로봇의 자율적인 행동에 관한 윤리적 문제, iii) AI 및 로봇 자체의 도덕적, 존재론적 지위 문제로 나누어 볼 수 있다. 우선, AI로 설계, 이용과 관련되는 알고리듬의 투명성, 공정성 문제는 결국 인간 개발자나 운영자의 윤리와 직결된다. '트롤리 딜레마'*11와 유사한 자율주행차의 윤리적 딜레마도 사실 인간의 윤리적 딜레마에 해당한다. 주행 중인 차량이 갑작스런 사고 발생 상황 하에서 인공지능의 알고리듬이 핸들을 꺾어 5명의 행인을 치거나 아니면 그대로 직진해 상대편 차량과 본인 차량 운전자 2명만 다치는 선택 중 하나를 택해야 하는 딜레마적 상황은 결국 알고리듬 설계자의 편향에 따라 좌우되는 윤리적 문제로 비화될 수 있다. 때문에 EU의회 등은 개발자, 설계자, 운영자, 이용자들을 위한 규범 체계인 윤리 가이드라인 제정을 시급한 과제로 요청하고 있다.

한편, '강한 AI '는 프로그래머가 모든 과정을 완벽하게 수행하더라도 학습 능력을 얻은 기계가 인간이 예측할 수 없는 결과를 초래할 수 있다는 딜레마로 이어진다. 따라서, AI 가 정확하게 작동하는지 여부를 떠나 로봇이 어떠한 규범적 체계에 근거해, 어떤 목적을 구체적으로 달성하려는지를 우선적으로 검증할 수 있어야 한다.

AI 및 로봇의 인격 주체성에 관한 문제는 미래 사회의 중요한 법철학적 쟁점이 될 것으로 전망된다. 세계적인 베스트셀러 '초지능(Superintelligence)'의 저자인 옥스퍼드대 닉 보스트롬(Nick Bostrom) 교수는 AI 시스템이 도덕적 지위를 갖기 위해서는 지각(Sentience)과 인격(Sapience) 요소를 갖춰야 한다고 설명한다. 인간과 짐승의 차이를 거론할 때, 동물은 인격을 갖추지 못하고 있기 때문에 인간이 좀 더 높은 도덕적 지위를 차지한다고 말할 수 있다. 보스트롬 교수는 두 존재가 동일한 기능과 동일한 의식 경험을 가지고 있고 그들의 성격만 다를 경우(Principle of Substrate Non-Discrimination), 또는 두 존재가 동일한 기능과 동일한 의식을 가지고 있고 그들의 근원만 다를 경우(Principle of Ontogeny Non-Discrimination), 이들은 동일한 도덕적 지위(Moral Status)를 갖는다고 설명한다. 즉, 인공지능이 두 가지 조건 중 하나에 부응하면 인간과 동일한 지위를 갖는다는 설명이다.

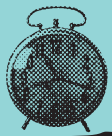
포스트휴먼(Post-human) 기술 및 윤리의 학제간 연구단체로 유명한 IEET(Institute for Ethics and Emerging Technologies)는 '비인간적 존재의 인격성'에 관한 논의를 학제적 연구와 실험을 통해 지속적으로 추진해오고 있다. 이처럼 인공지능 로봇, 사물인터넷 등 초연결사회의 도래가 예견됨에 따라 인간 이외의 사물이나 기계의 융복합적 권리, 책임 문제와 관련해 '인간을 넘어선 인격성(Personhood Beyond Human)'을 탐색하는 법철학적 연구는 갈수록 더욱 다각적으로 이뤄질 전망이다.

***10** 참고 | Peter Asaro(2015), 'Regulating Robots: Approaches to Developing Robot Policy and Technology' Presentation at: WeRobot 2015, University of Washington, April 10, 2015. http://www.werobot2015.org/wp-content/uploads/2015/04/Asaro_Regulating_Robots_WeRobot_ ***11** 참고 | 트롤리 딜레마(trolley problem)는 영국 철학자 필리파 푸트(Philippa R. Foot)가 처음 제안한 윤리학적 딜레마 문제이다. '100km로 주행 중인 기차(trolley)가 바로 앞에 인부 5명이 있는 것을 발견했지만 급정거를 하기에는 너무 가까운 거리인 반면 비상 철로에는 1명의 인부만이 철로를 청소하고 있다. 이럴 때 당신은 어떤 선택을 할 것인가'라는 딜레마적 상황을 제시하고 어떤 선택을 내릴 것인지 질문한다.

마치며

AI 에 대한 규제는 결국 AI 를 어떻게 바라볼 것인가에 따라 좌우된다. AI 가 인류의 존재까지 위협할 것이라 보는 시각은 정부의 적극적인 개입과 법제화를 요구하는 반면, AI가 가져올 변화들이 감내할 만한 수준이라 보는 시각은 가능한 최소한의 규제를 원할 것이다. 이처럼 AI 를 두고 엇갈리는 두 시선은 과소 또는 과도 규제를 둘러싼 논쟁을 끊임없이 초래할지 모른다. 다만 한가지 확실한 사실은 AI가 인류의 삶을 과거와는 확연히 다른 양상으로 바꿀 가능성이 더욱 높아졌다는 점이다. 인간을 닮은 AI 와 로봇을 어떻게 규율하느냐 문제는 인류가 그동안 접해보지 못한 새로운 규제 패러다임으로의 대전환을 요구하고 있다. 결국 미래를 어떻게 설계하고 그려갈 것인가는 인공지능이 아닌 인간의 의지에 달려 있는 셈이다.

로봇 윤리의 변천사.



글 | 김대원 ive.kim@kakaocorp.com

10년간 신문기자로 일했다. 기업 가치를 분석하고 거부들의 돈이 어디로 이동하는 지를 살폈다. 어릴 때는 장난감 로봇과 친하지 않았다. 박사 논문 주제로 로봇 저널리즘을 택하면서, 인공지능을 공부하기 시작했다. 새로운 기술과 미디어의 접목에 대해 논문 쓰는 취미를 갖고 있다.

사람을 대신할 인공지능(AI)을 마주하는 우리는 삶이 보다 편해지고, 윤택해질 미래를 그린다. 그러나, AI와 공존하게 될 미래가 마냥 아름답게만 그려지지 않는다. 디스토피아 시나리오는 인간을 위해 만들어진 기계가 외려 인간을 해칠 수 있지 않을지 염려한다. 인간을 위해 그리고 인간에 의해, 만들어지는 존재가 인간을 위협하지 않도록 인간은 부정적 개연성을 줄이고 싶어한다. 이러한 맥락에서 로봇 혹은 AI의 기술적 발달과 나란히 논의되는 주제가 윤리다. '윤리'라는 말로 포장되긴 하지만, 로봇 '윤리'[*1]는 인간을 이롭게 하려는 본래의 목적에 부합하기 위해 로봇이 지켜야 할 '준칙'이다. 이를 전제로 로봇의 윤리에는 인간과의 공존, 로봇 개발자와 이용자의 책임, 프라이버시 보호, AI 기반 무기 경쟁 지양, 초지능의 발전 방향 제시 등이 포함된다. 로봇 윤리의 고전인 '아시모프의 3원칙'부터 2017년 1월에 나온 최신 '아실로마 원칙'까지 관련 논의를 살펴보자.

***1** 참고 | 로봇윤리(roboethics)는 2002년 로봇공학자인 지안마르코 베루지오(Gianmarco Veruggio)에 의해 제안됐다.
2004년 이탈리아에서 열린 1회 국제로봇윤리 심포지엄에서 공식적으로 처음 활용됐다.

시즌 1 : 로봇의 책무만을 강조하던 시대

1942년, 아시모프의 로봇 3원칙

가장 널리 알려진 로봇 윤리 원칙은 아이작 아시모프(Issac Asimov)의 로봇 3원칙(The Three Laws of Robotics)이다.*2 3원칙, 혹은 아시모프의 원칙은 1942년 발간된 그의 단편 소설 '탑돌이*3(Runaround)'에서 제안됐다. 세 가지 원칙은 다음과 같다.

[아시모프의 로봇 3원칙]

1원칙 | A robot may not injure a human being or, through inaction, allow a human being to come to harm(로봇은 인간에게 해를 끼치거나, 아무런 행동도 하지 않음으로써 인간에게 해가 가도록 해서는 안 된다).

2원칙 | A robot must obey the orders given it by human beings except where such orders would conflict with the First Law(로봇은 인간의 명령에 복종해야 한다. 단 명령이 첫 번째 원칙에 위배될 때는 예외로 한다).

3원칙 | A robot must protect its own existence as long as such protection does not conflict with the First or Second Laws(로봇은 자신을 보호해야 한다. 단 첫 번째와 두 번째 원칙과 위배될 때는 예외로 한다)

아시모프의 3원칙은 로봇은 인간의 후생(厚生)을 위해 존재하며, 인간에 의해 인간을 해하는 방향으로 활용될 가능성을 경계한다. 로봇은 자율적으로 행동해야 하며, 스스로 자체의 내구성을 유지하고 보호해야 한다고 규정하고 있다. 다만, 자신을 위한 자위적(自衛的) 행동은 인간의 후생을 위한 명령보다 후순위다. 이같은 로봇의 딜레마는 "무엇이 인간을 위한 것인가?"라는 문제다. 인간이 원하는 바를 달성시켜 주기 위한 행동이 인간에게 해가 될 때, 이를 수행하는 것이 옳은 지 아니면 행동을 하지 않는 것이 바람직한지, 로봇은 혼돈에 빠질 수 있다. 로봇에 내재되는 프로그램이 어느 쪽에 무게를 두느냐에 따라, 로봇은 인간을 물리적으로 해치는 주체가 될 수 있다.

로봇을 소재로 한 소설, 영화 뿐만 아니라, 로봇 윤리와 유관 정책 연구에도 활용되고 있다. 로봇 3원칙의 현실 적용에 따른 로봇 윤리 문제에 대한 실질적 고민을 경험해 보고 싶은 사람은 영화화된 아시모프의 작품인 '아이, 로봇(I, Robot)'과 '바이센테니얼 맨(Bicentennial Man)' 감상을 권한다.

1942~

1985년, 아시모프의 로봇 0원칙

아시모프는 1985년 단편 소설인 '로봇과 제국(Robots and Empire)'에서 로봇 0원칙을 추가 제안했다. 로봇 0원칙은 다음과 같다. 0. 'A robot may not harm humanity, or, by inaction, allow humanity to come to harm'(로봇은 인류에게 해를 가하거나, 행동을 하지 않음으로써 인류에게 해가 가도록 해서는 안 된다).

로봇 3원칙의 첫 번째 원칙과 다를 바 없어 보이는 이 문구의 핵심적 차이는 '인류'다. 대동소이해 보이지만, 0원칙이 함의하는 바는 크다. 우선, 아시모프는 첫 번째 원칙의 상위 판단 기준으로 삼기 위해, 새로운 원칙에 '1'보다 앞선 '0'의 숫자를 부여했다. 두 원칙이 충돌할 때는 0원칙에 의해 로봇이 행동해야 함을 의미한다. '인간(Human)'을 '인류(Humanity)'로 전환은 공리주의적 발상에 기인한 것으로 해석된다. 아시모프는 한 인간보다는 인류의 이익을 지키는 방향으로 로봇의 행동이 결정되는 것이 합리적이라고 판단한 것이다.*4 0원칙에는 인류 멸망의 가능성을 명확하게 감지하면, 로봇이 한 개인의 이익 보다는 인류의 생존을 위한 결정이 바람직하다는 판단이 전제된 것이다.*5

1985~

***2** 참고 | 아시모프는 러시아 태생의 생화학자이며, 과학 저널리스트이자, 소설가였다. 아시모프는 과학과 대중을 잇는 매개자로서의 역할을 충실히 한 인물로 평가된다. 그가 쓴 과학소설은 500여 편에 이를 뿐만 아니라, 다양한 과학을 일반인에게 설명하는 대중서도 다수 집필했다. 로봇을 주제로 쓴 아시모프의 공상과학(SF) 소설 25편은 현대 로봇공학 발전에 기여했다고 평가받고 있다. "로봇 공학"이란 용어가 아시모프가 만들어 낸 신조어다. ***3** 설명 | 국내 번역서의 제목이다. ***4** 논문 | Clarke, R. (1993). Asimov's laws of robotics: implications for information technology-Part I. Computer, 26(12), 53-61. ***5** 참고 | 0원칙의 또 다른 해석도 존재한다. 색다른 해석은 외계인의 존재, 그리고 그것의 지구 공습을 상정한 것이다. 이러한 생각 하에서 0원칙은 외계인으로부터 인류가 공격을 받게 될 때, 로봇에게 이를 방어할 의무를 한 것이라고 해석됐다. 돌연변이 인간이나 바이러스의 위협도 인류를 위해 로봇이 막아야 할 대상으로 간주되기도 한다.

시즌 2 : 인간과의 공존, 인간의 책임

2004년, 일본 후쿠오카 세계 로봇 선언

2000년대 중반 이후의 로봇의 활용 분야가 다양해지고, 사람과 로봇 간의 접촉 빈도가 증가하면서, 사람과 로봇 간의 상호 작용에 대한 관심이 높아졌다.*6 이러한 흐름 하에서 2004년 발표된 로봇 윤리 원칙이 일본 후쿠오카 세계 로봇 선언이다. 후쿠오카는 일본에서는 로봇의 상징적인 지역이었다. 당시 일본에서 생산되는 로봇 생산 규모는 6조4000여억 원에 달했는데, 이 중 20%가 후쿠오카에서 생산된다. 일본 정부는 후쿠오카를 일본의 '로봇 특구'로 지정했다. 후쿠오카에서의 세계 로봇 선언 발표는 로봇 분야에서의 위상을 강화하겠다는 일본의 의지가 반영된 것이었다.*7 2004년은 로봇윤리를 주제로 한 제1회 국제로봇윤리 심포지엄이 이탈리아에서 열린 해이기도 하다.*8

[후쿠오카 세계 로봇 선언]

1. 차세대 로봇은 인간과 공존하는 파트너가 될 것이다.
2. 차세대 로봇은 인간을 육체적으로 그리고 정신적으로 보조할 것이다.
3. 차세대 로봇은 안전하고 평화로운 사회 구현에 기여할 것이다.

후쿠오카 선언의 핵심은 로봇과 인간 간의 공존이다. 공존한다는 점을 강조하면서도, 여전히 로봇이 인간을 해칠 가능성은 막아야 한다는 '안전 관리' 의지가 담겨 있다. 로봇이 발전할수록 위협에 대한 공포를 심화시키는 경향이 있다. 사실 로봇 혹은 AI를 어떻게 관리해야 하는 지에 대한 논의는 로봇 개발이 만들어낼 경제적 가치에 비해서는 상대적으로 소외되어 왔다.*9 로봇 그 자체의 기술적 발전 외에, 로봇이 초래할 역기능에 대한 관리, 그리고 그것을 위한 고찰의 필요성에 대한 문제 제기는 2000년 대 중반 부터 활성화 됐다. 그리고 로봇 자체 외, 로봇을 고안한 설계자가 책임 소재지(Locus)로서 다뤄지기 시작했다. 로봇 윤리의 수행 주체에 대한 논의가 변화된 것이다.

2004~

2006년, 유럽로봇연구연합(EURON)의 로봇윤리 로드맵

EURON은 유럽로봇연구연합(The European Robotics Research Network)에서 각 글자의 머릿말을 따서 만들 말이다. 2000년 유럽 연합(EU, 당시 EC) 산하의 기구로 만들어졌다. 유럽 로봇 연구 네트워크란 우리말로 해석될 수 있는 이 조직의 목적은 당시로부터 향후 20년 후를 위해 로봇에 내재된 기회를 명확하게 하고, 로봇 기술 발전을 통해 활용하는 것이다.[*10] EURON은 2003년부터 3년 간 로봇의 윤리 문제를 다루기 위한 로드맵(Road Map)을 설계했다. 인간과 로봇을 연구하는 연구자 50여 명이 참여, 로봇 개발 과정에서의 주된 윤리적 쟁점에 대한 구조적 평가(Systematic Assessment)를 시도했다. EURON의 로봇윤리 로드맵은 이전과 달리 '로봇이 어떠해야 한다'라는 원칙이 아닌 로봇을 만드는 사람을 대상으로 가이드라인을 제공하고자 했다. EURON이 로봇 윤리와 선행되는 원칙으로 제시한 항목은 아래와 같다.

[로봇윤리에 선행되는 원칙들]

1. 인간의 존엄과 인간의 권리
2. 평등, 정의, 형평
3. 편익과 손해
4. 문화적 다양성을 위한 존중
5. 차별과 낙인화의 금지
6. 자율성과 개인의 책무성
7. 주지된 동의
8. 프라이버시
9. 기밀성
10. 연대와 협동
11. 사회적 책무
12. 이익의 공유
13. 지구 상의 생물에 대한 책무

로봇이 아닌, 로봇 개발자 그리고 운영자에 윤리적 책임을 집중해야 하는 이유에 대한 여러 논의가 진행된 가운데 과학철학자인 닉 보스트롬(Nick Bostrom)은 신기술과 인류에 대해 통찰력을 보여주고 있다. 그는 인공지능 창조에 대한 윤리 법칙이란 그의 글을 통해[*11], 다음과 같이 강조했다.

"창조자는 그들이 만든 산물(Progeny)의 행위에 대한 책임이 있다. 만일, 창조자가 도덕적 지위를 갖춘 창조물을 만든다면, 창조자와 창조물은 창조물의 행동에 대해 공동 책임을 져야 한다."

2006~

2007년, 산자부의 로봇윤리헌장

국내에서는 산업자원부가 2007년 공개한 '로봇윤리헌장' 초안에서 로봇의 행동 책임 범위를 로봇을 활용하는 주체까지 포함시켰다. 로봇의 윤리, 제조자의 윤리에 이어, 사용자의 윤리까지 로봇 윤리 범주에 포함시킨 것이다. 로봇, 제조자, 사용자까지 포괄한 로봇 윤리 규정을 제정한 것은 세계 최초의 일이었다. 이 윤리 헌정을 만드는 일에는 정부 관계자, 로봇 공학 교수, 심리학 전문가, 의사 등 12명이 참여했다. 다만 산업 현장의 이해 관계자 의견까지 포괄한 뒤, 2007년 말쯤 공식 발표될 것이란 당시 전망과 달리, 한국의 로봇 윤리 헌정은 아직도 미생(未生)으로 남아 있다.

[산업자원부 로봇윤리헌장(초안)]

1장(목표) | 로봇윤리헌장의 목표는 인간과 로봇의 공존공영을 위해 인간중심의 윤리규범을 확인하는 데 있다.

2장(인간, 로봇의 공동원칙) | 인간과 로봇은 상호간 생명의 존엄성과 정보, 공학적 윤리를 지켜야 한다.

3장(인간 윤리) | 인간은 로봇을 제조하고 사용할 때 항상 선한 방법으로 판단하고 결정해야 한다.

4장(로봇 윤리) | 로봇은 인간의 명령에 순종하는 친구 · 도우미 · 동반자로서 인간을 다치게 해서는 안 된다.

5장(제조자 윤리) | 로봇 제조자는 인간의 존엄성을 지키는 로봇을 제조하고 로봇 재활용, 정보보호 의무를 진다.

6장(사용자 윤리) | 로봇 사용자는 로봇을 인간의 친구로 존중해야 하며 불법개조나 로봇남용을 금한다.

7장(실행의 약속) | 정부와 지자체는 헌장의 정신을 구현하기 위해 유효한 조치를 시행해야 한다.

2007~

***6** 논문 | 이원형, 박정우, 김우현, 이희승, & 정명진. (2014). 사람과 로봇의 사회적 상호작용을 위한 로봇의 가치효용성 기반 동기-감정 생성 모델. 제어로봇시스템학회 논문지, 20(5), 503-512. ***7** 참고 | http://ouic.kaist.ac.kr/news04/articles/do_print/tableid/news/category/7/page/24/id/715 ***8** 참고 | 당시 이러한 시대적 흐름 속에 국내에서도 로봇 엑스포를 개최하자는 논의가 제기됐다. ***9** 논문 | 주일엽. (2011). 지능형 로봇에 대한 안전관리 발전방안. <한국경호경비학회>, 26호, 89-119. ***10** 참고 | http://www.roboethics.org/atelier2006/docs/ROBOETHICS%20ROADMAP%20Rel2.1.1.pdf ***11** 논문 | Bostrom, N. (2007). Ethical principles in the creation of artificial minds. Linguistic and Philosophical Investigations, 6, 183-184.

시즌 3 : 프라이버시와 투명성의 강조

2010년, 영국의 공학과 물리과학 연구위원회의 로봇 원칙

2000년대 들어서 국제적 차원의 논의가 여러 차례 일어났지만 아직도 로봇이나 인공지능은 먼 미래의 이야기로 여겨졌다. 전문가들은 여전히 아시모프의 3원칙 수준에 머물러 있었다. 선행해서 로봇 윤리를 논하던 연구자들의 고민은 학계에 제대로 뿌리내리지 못했다. 2010년 9월, 영국의 기술, 산업, 예술, 법의 전문가들이 자체적인 로봇 원칙을 마련하고자 모였다. 이른바 '공학과 물리과학 연구위원회'(Engineering and Physical Sciences Research Council, EPSRC). 그들의 문제 의식은 "실험실(Lab)에서 집으로, 산업 현장으로 나온 로봇을 위한 원칙이 필요한데 아직도 로봇에 대한 논의는 공상 과학 소설이나 영화에 머물러 있는" 현실에서 출발했다.

소설 세계가 아닌 현실 영역 속 로봇의 관리를 위해 EPSRC는 로봇 원칙 5을 제안한다.[*12] EPSRC가 다른 원칙과 다른 지점은 프리이버시에 대한 언급이다. EPSRC는 로봇에 의해, 사람의 정보가 일거수 일투족 쌓일 수 있다는 지점을 경계했다. EPSRC는 원치 않은 개인의 일상이 정보 형태로 축적되어 공개되는 상황을 우려했다. 데이터가 프라이버시와 연계되어 로봇 혹은 AI의 관리를 위해 고려되어야 할 요인으로 언급된 것은 당시로서는 이례적이었다. 사실, AI 시대에 빅데이터의 활용 구상과 범위는 AI 품질을 결정하는 핵심적 요소이다. 현재 AI 핵심은 딥러닝의 핵심인 기계학습의 경쟁력은 기계가 학습할 수 있는 양질의 데이터 규모에 의해 좌우되기 때문이다.

[EPSRC의 로봇 윤리]

1. 로봇은 단 국가안보를 위한 경우는 제외하고, 무기로서 설계되어서는 안된다.
2. 로봇은 사생활 보호 뿐 아니라 프라이버시를 포함한 현존하는 법규범에 부합되도록 설계되고 운용되어야 한다.
3. 로봇은 안전과 보안 보장에 적합하도록 설계되어야 한다.
4. 로봇은 사람에게 이에 대한 착각이나 환상을 불러 일으키도록 설계되고 사용되어서는 안된다.
5. 모든 로봇에 대해 법적 책임을 지는 사람이 명확하게 명시되어야 한다.

2010~

2016년, 일본의 AI R&D 가이드라인과 사티아 나델라의 인공지능 규칙

AI 혹은 로봇 윤리에서 프라이버시 이슈는 주요 과제로서 지속적으로 논의되고 있다. 일본 정부가 발표한 AI R&D 가이드라인, 그리고 사티아 나델라(Satya Narayana Nadella) 마이크로소프트(MS) CEO가 발표한 AI 규칙에도 프라이버시 문제는 주요 사안으로 등장했다.

[일본의 AI R&D 가이드라인]

1. **투명성** | AI 네트워크 시스템을 설명하고 검증 할 수 있는 능력 확보
2. **이용자 지원** | AI 네트워크 시스템이 이용자를 지원하고 이용자에게 선택의 기회를 적절히 제공
3. **통제** | 인간에 의한 AI 네트워크 시스템의 통제 가능성 보장
4. **보안** | AI 네트워크 시스템의 견고성 및 신뢰성 확보
5. **안전** | AI 네트워크 시스템이 사용자 및 제 3 자의 생명 또는 신체에 위험을 초래하지 않도록 고려
6. **프라이버시** | AI 네트워크 시스템이 사용자 및 제 3 자의 개인 정보를 침해하지 않도록 고려
7. **윤리** | 네트워크화 될 인공 지능의 연구 개발을 수행함에 있어 인간의 존엄성과 개인의 자율성 존중
8. **책임** | AI 연구원 또는 개발자와 네트워크를 통해 사용자와 같은 이해 관계자에 대한 책임성 부여

일본의 AI R&D 가이드라인 8개항은 2016년 4월 가가와 현에서 열린 주요 7개국(G7) 정보통신장관 회의에서 제안됐다. 일본 정부는 "사생활 보호 또는 보안 규정을 다루는 경제협력개발기구(OECD) 가이드라인을 고려하면서, 네트워크화 될 AI에 대비하자"고 역설했다. 네트워크의 개념을 강조하면서, 부각된 가치는 '투명성'이었다. 일본 정부는 공유의 매개가 될 네트워크 시스템을 설명하고, 검증할 수 있는 능력의 확보를 강조했다.[*13]

[사티아 나델라의 인공지능 규칙]

1. 인공지능은 인간을 돕기 위해 개발되어야 한다.
2. 인공지능은 투명해야 한다.
3. 인공지능은 인간의 존엄성을 파괴하지 않으면서 효율을 극대화해야한다.
4. 인공지능은 개인정보 보호를 위해 설계 되어야한다.
5. 인공지능은 의도하지 않은 피해를 인간이 복구할 수 있도록(인간에 대한) 알고리즘 차원의 설명 책임을 지닌다.
6. 인공지능은 차별 또는 편견을 방지해야 한다.

나델라 MS CEO는 2016년 6월 미국의 온라인 매체인 슬레이트(Slate)에 칼럼 형식으로 자신이 생각하는 AI 규칙 6가지를 제시했다. 나델라의 원칙에서는 '편견의 방지'가 AI의 책무로 제시된 것이 특징이다. 이는 AI의 토대가 되는 알고리즘과 관련하여, 그 안에 내재될 수 있는 편향성이 점차 사회적 이슈로 부상되는 현실이 반영된 결과로 풀이된다. 이밖에, AI를 기존과는 완전히 다른 '플랫폼'으로 규정한 접근 역시도 주목할 만하다.

아마존의 알렉사(Alexa)의 등장 이후, AI 플랫폼 현상은 현실화 되는 양상이다. 아마존은 음성인식 기술의 오픈 API를 정책을 펴면서, 최신 IT서비스를 아마존의 생태계에 가두는 효과를 발휘했다. 아마존 알렉사는 2017년 2월 열린 소비자가전전시회(Consumer Eletronics Show, CES)에서 부스 하나 설치하지 않고도 행사 내내 주목받았다. IT업체들이 자신들의 첨단 이미지를 구축하기 위해 아마존의 음성인식 기술을 탑재한 데 따른 결과였다. 미디어 레퍼토리(Media Repertoire)[*14] 측면에서도 아마존의 음성인식 기술은 모바일을 대체할 '미디어'로서의 가능성을 보이고 있다. 미국 알렉사에 친숙한 미국의 초등학생들은 과제를 하기 위해, 모바일로 구글앱에 접속해 검색어를 입력하기 보다는 알렉사를 향해 말하고 있다. "알렉사, 미국의 45대 대통령이 누구야?".

2016~

***12** 참고 | https://www.epsrc.ac.uk/research/ourportfolio/themes/engineering/activities/principlesofrobotics/s ***13** 참고 | AI R&D 원칙을 제시한 당위성을 강조하기 위해, 일본이 설명한 1차원 이유는 AI가 인간을 해하거나, 악의적 목적에 의해 활용되거나, 혹은 통제 불능 상태에 빠져 사고를 발생하는 가능성을 미연에 방지하자는 것이었다. 그러나, 일본이 AI 활용에 대한 가이드라인 제정에 선도적인 움직임을 보인 궁극적인 목적은 AI의 안전적 이용 측면에서의 세계적 주도권을 잡기 위한 포석을 놓기 위한 것으로 해석되기도 한다. 2016년 일본 총무성의 전문가 회의는 2045년에 AI로 인해 창출될 일본 국내 경제효과를 121조 엔(1,222조 원)으로 예측했다. ***14** 참고 | 이용자가 미디어 소비 목적에서 즐겨 이용하는 미디어의 조합

시즌 4 : AI 무기 경쟁 경계와 초지능에 대한 고려

2017년, 아실로마의 AI 23원칙

아실로마 원칙(Asilomar AI Principles)은 삶의 미래 연구소(Future of Life Institute, FLI)*15가 2017년 1월 개최한 컨퍼런스인 '유익한 AI, 2017' (Beneficial AI 2017)에서 발표됐다. 이 컨퍼런스를 주관한 FLI는 스카이프 창업자인 얀 탈린(Jaan Tallinn) 등 5인이 주도로 설립한 민간 단체이다. FLI는 기술이 인간을 번성하게 하는 동시에 자멸하게 만드는 잠재력을 가지고 있으니, 다른 미래를 만들기 위해 노력하자고 강조하는 단체이다. 바이오 기술, 핵, 기후, 그리고 AI를 주된 논의 과제로서 삼고 있다.

FLI는 2015년에도 AI 활용한 킬러 로봇 개발의 위험성을 경고하는 서한을 만들어 발표하기도 했다. 이 서한에는 전 세계 1000명의 전문가가 함께 했다. FLI는 2017년 1월 AI의 건전한 발전을 위해 또 한 번의 컨퍼런스를 열었고, 그것이 앞서 설명한 '유익한 AI, 2017'이었다. 이 행사가 미국 캘리포니아 아실로마에서 진행되어, 발표된 AI 원칙의 이름이 '아실로마 원칙'이 된 것이다. 아실로마 원칙은 연구 관련 쟁점, 윤리와 가치, 장기적 이슈 등 세 가지 범주로 나뉜다. 아실로마 원칙에서는 기존의 원칙에서 논의된 사항 외에 이공지능을 활용한 무기 경쟁이 지양되어야 함을 강조했다. 초지능(Superintelligence)*16이라는 인간의 두뇌를 뛰어넘은 지능을 가진 존재에 대한 발전 방향이 제시된 점도 아실로마 원칙의 특징이다.

[아실로마의 AI 원칙]

연구 관련 쟁점

1. **연구목표** | 인공지능 연구의 목표는 방향성이 없는 지능이 아니라 인간에게 유용한 지능을 개발해야 한다.

2. **연구비 지원** | 인공지능에 대한 투자에는 컴퓨터 과학, 경제, 법, 윤리, 사회 연구와 같은 어려운 문제들을 포함해 유용한 이용을 보장하기 위한 연구비 지원이 수반되어야 한다.
   - 미래의 인공지능 시스템을 어떻게 강력하게 만들어 오작동이나 해킹을 할 수 없이 만들 수 있는가?
   - 인적 자원과 목적을 유지하면서 자동화를 통해 인간의 번영을 성장시킬 수 있는가?
   - 인공지능과 보조를 맞추고 인공지능과 관련된 위험을 관리하기 위해 보다 공정하고 효율적으로 법률 시스템을 업데이트 할 수 있는 방법은 무엇인가?
   - 인공지능은 어떤 가치체계를 갖추어야하며, 인공지능이 가져야 할 합법적이고 윤리적인 지위는 무엇인가?

3. **과학정책 연계** | 인공지능 연구자와 정책 입안자 사이에 건설적이고 건전한 교류가 있어야 한다.

4. **연구 문화** | 인공지능 연구자와 개발자 사이에 협력과 신뢰, 투명성의 문화가 조성되어야 한다.

5. **경쟁 회피** | 인공지능 시스템을 개발하는 팀들은 안전기준을 부실하게 만드는 개발을 피하기 위해 적극적으로 협력해야 한다.

윤리와 가치

6. **안전** | 인공지능 시스템은 작동 수명 전체에 걸쳐 안전해야하며, 적용가능하고 가능한 한 검증할 수 있어야 한다.

7. **실패의 투명성** | 인공지능 시스템이 손상을 일으킬 경우 그 이유를 확인할 수 있어야 한다.

8. **사법적 투명성** | 사법제도 의사결정에 자율시스템이 개입된다면 권위 있는 기관이 감사하고 이에 만족스러운 설명을 제공해야 한다.

9. **책임성** | 첨단 인공지능 시스템의 설계자와 구축자는 인공 지능의 사용, 오용 및 행동의 도덕적 영향을 미치는 이해 관계자이며, 그에 따른 책임과 기회가 주어진다.

10. **가치 일치** | 고도로 자율적인 인공지능 시스템은 목표와 행동이 인간의 가치와 일치하도록 설계해야 한다.

11. **인간의 가치** | 인공지능 시스템은 인간의 존엄성, 권리, 자유 및 문화적 다양성의 이상에 적합하도록 설계되고 운용되어야 한다.

12. **개인정보 보호** | 인공지능 시스템이 사람들이 생성한 데이터를 분석하고 활용할 수 있는 권한을 부여 받으면, 사람들은 그 데이터에 접근하고 관리하며 통제할 수 있는 권리를 가져야 한다.

13. **자유와 프라이버시** | 인공지능을 개인정보에 적용하게 되더라도 사람들의 실제 또는 인지된 자유가 부당하게 축소되어서는 안된다.

14. **이익의 공유** | 인공지능 기술은 최대한 많은 사람들에게 혜택을 주고 힘을 부여해야 한다.

15. **번영의 공유** | 인공지능에 의해 만들어진 경제적 번영은 인류 모두의 혜택을 위해 널리 공유되어야 한다.

16. **인간통제** | 인간은 선택한 목표를 달성하기 위해 의사결정을 인공지능 시스템에 위임하는 방법과 위임 여부를 선택해야 한다.

17. **사회 전복 방지** | 고도의 AI 서비스를 통제함으로써 생기는 힘은 인류 발전을 위해 쓰여야 한다. 건강한 사회를 지향하고, 이를 지키려는 사회나 시민들의 절차를 존중하고 개선해야 한다.

18. **인공지능 무기 경쟁** | 치명적인 인공지능 무기의 군비 경쟁은 피해야 한다.

장기적 이슈

19. **역량 경고** | 합의된 여론이 없는 상태에서 미래 인공지능 역량의 한계에 대한 강력한 가설은 삼가야 한다.

20. **중요성** | 고도화된 인공지능은 지구상의 생명의 역사에 심각한 변화를 가져올 수 있다. 따라서 그에 상응한 관심과 자원을 계획하고 적절하게 관리해야 한다.

21. **위험성** | 인공지능 시스템이 초래하는 위험은 특히 실존적이거나 파괴에 관련된 것들이다. 앞으로 예상되는 영향에 맞는 계획을 세우고 위험을 완화하기 위한 노력을 해야 한다.

22. **자기 개선 순환** | 인공지능 시스템은 스스로 발전하거나 자가 복제할 수 있게 설계된다. 그러므로 양적으로나 질적으로 급격히 증가할 수 있기에 엄격한 안전 관리 및 제어 조치가 수반되어야 한다.

23. **공동의 선** | 초지능은 윤리적 이상을 널리 공유하는 방식으로 발전되어야 한다. 그리고 한 국가나 조직보다 모든 인류의 이익을 위해 개발되어야 한다.

2017~

***15** 참고 | https://futureoflife.org/ ***16** 참고 | 이 행사에는 에릭 슈미트 구글 회장, 앤드류 응 중국 바이두 최고과학자 등 AI와 관련된 전 세계 유수의 전문가가 참석했다. 일론 머스크 테슬라 CEO, '알파고' 아버지인 데미스 하사비스 딥마인드 대표, 미래학자 레이 커즈와일, 닉 보스트롬 옥스퍼드대 인류미래연구소 등 9인이 '초지능'에 대한 토론을 벌이기도 했다. 이 내용은 카카오 정책지원팀의 브런치에 정리되어 있다(https://brunch.co.kr/@kakao-it/49).

2017년 AI 컨퍼런스

최근 IT기업 뿐 아니라 다양한 분야의 기업들이 AI연구에 관심을 보이고 있다. 이와 관련된 다양한 행사와 학회들이 2017년에도 예정되어 있다. 최신 연구결과를 확인 할 수 있을 뿐 아니라, 다양한 분야의 기업들이 AI를 활용하는 사례들을 만나 볼 수 있다.

O'Reilly Summit (The O'Reilly Artificial Intelligence Conference)

O'Reilly는 프로그래밍 교육 교재를 제작 판매하는 기업 2016년 부터 시작한 컨퍼런스로, AI 연구자들과 AI 를 활용한 기업들 사례들, 강연들이 이루어진다.

Date & Location
6월26일~6월29일, 뉴욕
9월 17일~20일, 샌프란시스코

Information
https://conferences.oreilly.com/artificial-intelligence/ai-ca

CVPR (Computer Vision and Pattern Recognition)

IEEE 라는 글로벌 테크노놀로지 전문가 그룹 조직에서 주관하는 행사 1991년 부터 시작된 컨퍼런스. 컴퓨터 비전(지각)과 이미지처리등에 관한 최신 연구 결과를 확인할 수 있다.

Date & Location
7월22일~7월25일, 하와이

Information
http://cvpr2017.thecvf.com

The AI Summit

2014년도 부터 시작된 행사로 AI와 비즈니스의 만남을 주제로 진행되는 컨퍼런스 다양한 비즈니스 분야에서 AI가 실질적인 응용 사례들을 접할 수 있다.

Date & Location
5월 9일 런던
9월27일 샌프란시스코
12월6일 뉴욕

Information
https://theaisummit.com

ICML (International Conference on Machine Learning)

2017년은 34번째 행사로, 오랜 역사를 가지고 있는 머신러닝 관련 최대의 학회 모든 머신러닝 연구자들이 논문을 발표하기 원하는 학회로 NIPS에서는 주로 신경망과 관련된 논문이 주로 발표 된다면, ICML은 기계학습, 데이터마이닝 등과 관련된 논문이 중심이 된다.

Date & Location
8월6일 ~ 8월11일, 시드니

Information
https://2017.icml.cc

NIPS (Neural Information Processing Systems)

1987년 캘리포니아공과대학(Caltech)과 벨연구소가 기획해서 최초의 컨퍼런스가 열렸다. 당시 아웃사이더 였던 인공신경망(Artificial Neural Networks)연구자들을 위한 그들만의 학회에서 현재는 AI, 머신러닝, 통계 분야의 가장 많은 논문이 발표되는 학회가 되었다. 가장 최근에 열린 2016년 12월에서도 약 500여편의 논문이 발표 되었다.

Date & Location
12월4일 ~ 12월9일, 캘리포니아

Information
https://nips.cc

AI 팟캐스트

2012년 이후 부터는 IT관련 미디어에 AI가 종종 등장하고 있었다. 최근들어 기술적인 내용 뿐 아니라 AI의 사회적 효과 및 이슈들에 대해서 다루고 있는 미디어들도 늘어나고 있다. 아직 한국어로 제공되는 방송이 없어 아쉽지만, AI에 관한 최신 소식을 접하고 싶다면 다음의 Podcast를 추천한다.

The AI Podcast

Nvidia에서 제공하고 있는 Podcast로 Fortune Magazine 출신 기자가 진행하고 있다. The AI Podcast는 AI관련 최신 뉴스 및 연구자들 인터뷰를 통해 기술을 소개하고 있다. Nvidia는 GPU가 AI기술에 중요한 역할을 담당하면서 가장 빠르게 성장하고 있는 기업으로 최근 AI기술의 활용과 관련된 다양한 홍보를 진행 중에 있다.

difficulty | ★★

https://itun.es/kr/RkeUgb.c

Concerning AI | Artificial Intelligence

작가 2명이 진행하는 Podcast로, 다른 방송들과는 다르게 기술적인 내용 보다는 AI와 관련된 사회적 이슈들에 대해서 자유롭게 대화로 풀어간다. 영화속에 등장한 AI에 대한 진행자들의 생각, 인간과 대결하는 AI에 대한 이야기를 하고 있다. 배경 지식 없이 들을 수 있는 방송이다.

difficulty | ★

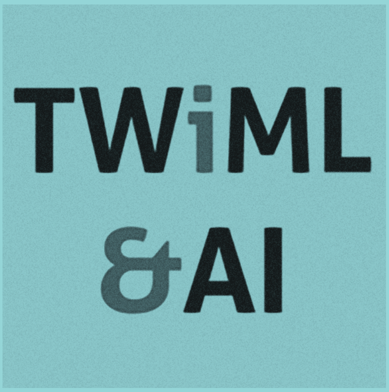
https://itun.es/kr/RJz69.c

Learning Machines 101

유타 대학의 컴퓨터 사이언스 교수가 진행하는 머신러닝에 대한 기초강의. 지금까지 약 60개 정도의 방송이 제공되고 있는데, 처음부터 듣게 되면, 이해하기 쉽게 머신러닝 및 AI에 대해서 공부 할 수 있다. 홈페이지를 찾아 보면, 강의에 대한 Script와 예시 이미지도 제공하고 있다.

difficulty | ★★★

https://itun.es/kr/FYRn1.c



This Week in Machine Learning & AI Podcast

IT 컨설팅 전문가들이 진행하는 Podcast. AI관련 이슈를 전문가의 인터뷰를 통해 설명하고 있다. AI 컨퍼런스에 참석한 연구자들과 현장 인터뷰를 통해 연구 성과를 들을 수 있다.

difficulty | ★★

https://itun.es/kr/l7wIcb.c

Talking Machines

하버드대 응용과학 교수가 진행하는 Podcast로 소개된 방송들 중 가장 난이도가 높은 편이다. 용어설명, Q&A, 인터뷰의 3가지 섹션으로 구성되어 있으며 NIPS, ICML등의 중요 컨퍼런스가 진행된 경우 관련 논문을 소개하기도 한다. 2016년 9월을 마지막으로 방송은 종료 되었지만, AI와 머신러닝에 대해서 공부 할 수 있는 방송이다.

difficulty | ★★★★

https://itun.es/kr/D0Y74.c

마치며

AI는 긴 겨울 'AI Winter'을 끝내고 빠르게 발전하고 있습니다. AI에 대한 논의도 봇물터지듯 쏟아집니다. 막연한 우려나 기대 대신 다양한 논쟁들을 풀어나갈 방향성을 고민해야 할 때입니다. 각 현안 별로 생산적 담론을 끌어내지 못한다면 다시 겨울을 맞이할 수도 있습니다.

카카오 AI 리포트는 AI와 관련된 다양한 내용을 꾸준히 소개하려고 합니다. 어려운 AI 기술과 개념들을 조금이라도 쉽게 풀어보거나, 다양한 이슈들을 점검해보며, 새로운 소식들을 전할 예정입니다. 단순 지식이나 정보 전달을 넘어 전문가들의 견해를 바탕으로 좀 더 깊이 있는 내용을 전달하려 합니다. AI가 공상과학이나 단편적인 이벤트로 화제를 모으는 것을 넘어, 현재와 미래를 이어주는 실질적인 기술로 이해되기를 기대합니다.