

#3

AI 기술의 최신 트렌드와 산업계 적용 방향성



글. 조성준 서울대학교 산업공학과 교수



인공지능(Artificial Intelligence, AI)은 사람처럼 학습하고 문제를 해결하는 컴퓨터 프로그램으로서 예측, 이상 탐지, 자연어 처리, 이미지 인식 등 다양한 분야에서 활용된다. AI는 작동 방식에 따라 분석형 AI와 생성형 AI로 나뉜다. 분석형 AI는 데이터를 수집하고 분석하여 패턴을 찾거나 yes/no 같은 판정을 하는데 사용된다. 이러한 기술은 예측, 추천, 문서 분류나 문서의 정서 분석 등의 분야에 활용된다. 생성형 AI는 새로운 콘텐츠를 만

들어내는 데 사용된다. 예를 들어, ChatGPT와 Stable Diffusion은 문서 요약, 보고서 작성, 작곡, 작사 및 이미지 디자인 등의 분야에 활용된다. AI의 정의는 목표를 제시한 것이고, 어떻게 만들 것인지에 대해서는 언급이 없다. 어떻게 만들든 사람 같으면 되는 것이다. 그래서 전문가에게 어떻게 하는지 물어보는 방법과 실제 전문가가 어떤 결정을 하는지 기록을 했다가 그 데이터를 열심히 보고 배우는 방법, 두 가지로 AI를 만드는 시도를 했다.

지식 기반 방법

먼저, 지식 기반 방법은 전문가의 지식을 명제로 표현하여 컴퓨터가 이를 이해하고 연역적으로 추론할 수 있도록 하는 방법이다. 이를 위해서는 지식 공학(Knowledge engineering)이 필요하다. 예를 들어 의료 진단에서는 전문의의 지식을 명제로 표현하여 컴퓨터가 환자의 증상과 검사 결과를 분석하여 질병을 진단할 수 있다. 또한 법률 분석에서는 전문 변호사의 판례와 법률 지식을 명제로 표현하여 컴퓨터가 유사한 사례를 분석하고 판결을 예측할 수 있다. 자연어 처리에서도 지식 기반 방법이 사용된다. 예를 들어, 자동 번역 시스템에서는 언어 간 문법 차이와 단어 의미를 이해하기 위해 언어학적 지식이 필요하다. 이러한 지식은 명제 형태로 표현되며, 컴퓨터가 이를 이해하고 번역할 수 있도록 한다.

지식 기반 방법의 한계

그러나 지식 기반 방법은 크게 세 가지 한계가 있다. 첫째, 작업이 어려워지면 명제 집합, 즉, 지식베이스(Knowledge Base)가 너무 커져서 구축하거나 관리하기가 아주 어려워진다. 둘째, 전문가가 여럿이면 각자의 주관에 따라 명제가 다르고, 심지어 동일인도 동일한 상황에서 비밀관성 때문에 다른 결론에 다다르게 되어 지식 베이스가 일관성 없이 구축될 수 있다. 셋째, 암묵적인 지식(implicit knowledge)을 명시적으로 표현하기 어렵다. 예를 들어, 의료 진단에서 전문가가 가진 경험과 직관을 표현하기 어렵고, 사람이 걸어갈 때 언제 어떤 근육을 수축하고 이완하는지 명시적으로 표현하는 것도 불가능하다. 즉, 잘하는 사람이 스스로 어떻게 잘하는지 명제로 표현하지 못하는 경우는 셀 수도 없이 많다.

데이터 기반 방법

데이터 기반 방법은 머신러닝(Machine Learning)이라고도 불리며, 데이터로부터 패턴을 학습하여 귀납적으로 예측하는 AI를 만드는 방법이다. 대표적인 머신러닝 알고리즘으로는 신경회로망(Neural Networks)과 초대형 신경회로망을 딥러닝(Deep Learning)으로 학습한 AI가 강력한 성능을 보이고 있다. 예를 들어 DNN(Deep Neural Network), CNN(Convolutional Neural Network), Transformer, BERT, GPT, LLaMA 등의 알고리즘이 있으며, 이들은 이미지나 자연어 처리 분야에서 높

은 성능을 보여주고 있다. 구체적인 사례로는 스팸 필터링, 이미지 인식, 음성 인식 등이 있다. 예를 들어 스팸 필터링에서는 이메일의 제목과 내용, 발신자 정보 등의 데이터를 학습하여 스팸 메일을 자동으로 걸러내는 AI를 만들 수 있다. 이미지 인식에서는 대량의 이미지 데이터를 학습하여 컴퓨터가 사물이나 인물 등을 자동으로 인식할 수 있도록 한다. 음성 인식에서는 음성 데이터를 학습하여 컴퓨터가 음성 명령을 이해하고 실행할 수 있도록 한다. 최근에 소셜미디어와 IoT(Internet of Things)의 발전으로 빅데이터가 등장하면서 데이터 기반 AI 구축 방법이 주목받고 있으며, CPU와 GPU 속도 향상으로 학습 시간이 단축되어 실용적인 방법이 되고 있다.

데이터 기반 방법의 한계

데이터 기반 방법에도 한계는 있다. 첫째, 데이터가 없으면 AI를 만들 수 없다. 둘째, 데이터로부터 학습한 것만을 알고 다른 정보를 활용하지 못한다. 셋째, 세상이 바뀌면 새로운 데이터를 수집하고 AI를 재학습해야 한다. 예를 들어, 카드 사기 검출 AI는 카드 사기 패턴이 계속해서 변화하기 때문에 새로운 데이터를 수집하고 AI를 재학습해야 한다. 또한 대출 심사 AI는 코로나 전과 코로나 중, 그리고 코로나 후가 다른 상황에서 다른 학습데이터로 학습되어야 한다.

AI 적용 가능 분야

AI를 어떤 비즈니스 분야에 적용할 수 있을까? 결론부터 말하자면, 적용 비즈니스 분야는 기업에서 하는 일 전부 다라고 할 수 있다. 제품/서비스 기획부터 시작하여, 제품/서비스 설계, 구매/유통/공급망, 제조/품질, 마케팅/프로모션/광고, 영업/CRM/AS, 재무/회계/투자, R&D, HRD 등이 포함된다. 이러한 영역에서 AI를 활용하여 소비자 파악, 설계오류 분류, 원자재가격 예측, 품질/불량예측, 제품 컨셉 디자인, 프로모션, 수요 예측, 기술동향 파악, 인사평가 등의 작업을 수행할 수 있다(그림 1).

이러한 비즈니스 가치를 창출하기 위해서는 4단계로 일을 진행해야 한다. 먼저, 1단계는 기획이다. 어떤 업무에 AI를 적용할지, chatGPT 같은 외부 AI를 사용할 것인지 자체 개발할 것인지, 자체 개발한다면 어떤 데이터로 학습할 것인지, 장단점을 고려하여 결정해야 한다. 2단계는 외부 AI 가운데 어떤 것을 구매할 것인지 결정하거나, 내부, 외부 데이터를 확보하여 AI를

그림 1. 경영 기능 별 AI 적용 기회

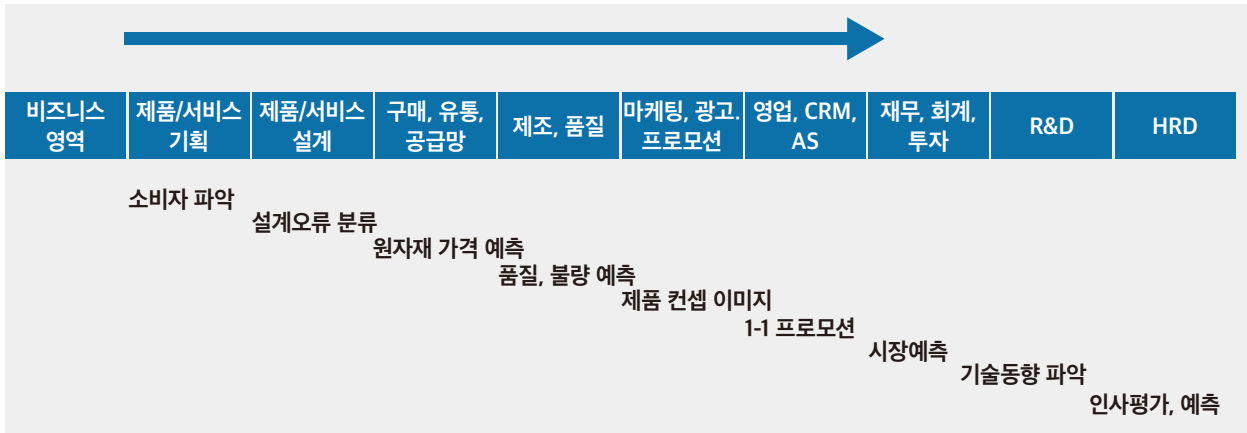
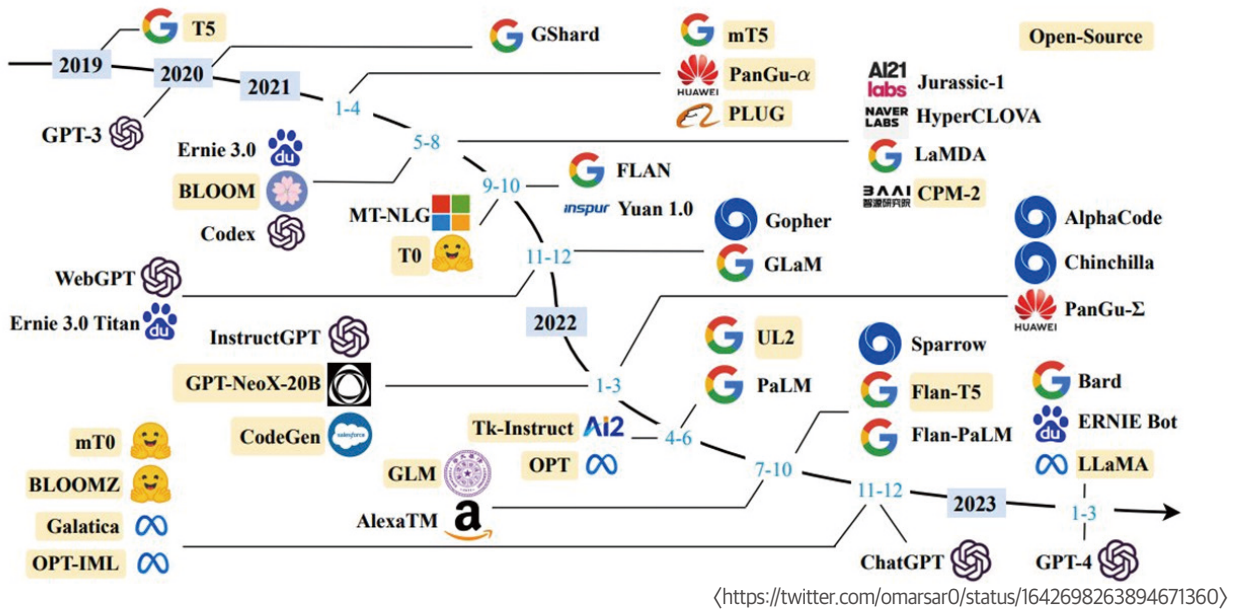


그림 2. 2019~2023.3 사이에 발표된 AI. 이후에 Dolly 2.0 (2023.4) 와 MPT-7B (2023년 5월) 출시



개발하는 단계다. 3단계는 (구매한 또는 개발한) AI를 현업의 사결정에 적용하기 전 평가하는 단계다. 4단계는 AI 적용 및 적용 후 평가단계이다.

단계별 과업 수행자

각 단계는 기업 내에서 누가 해야 할까? 먼저, 1단계 기획은 현업 의사결정자가 수행한다. 왜냐하면 어떤 일에 AI를 적용할지 판단하는 건 현업 의사결정자가 제일 잘 알 수 있기 때문이

다. 2단계 구매는 현업 의사결정자가 AI 전문가의 도움을 받을 수 있고, 자체 개발은 AI 전문가가 주도한다. 다만, 이때도 학습데이터 확보는 현업 의사결정자가 적극 도와야 한다. 3단계도 현업 의사결정자들이 AI의 적용 가능성과 효과를 평가한다. 마지막으로 구매/개발된 AI를 실제로 적용하고 그 효과를 평가하는 것 역시 현업 의사결정자의 몫이다. 따라서, 비즈니스 가치 창출을 위해서는 기획부터 평가까지 현업 의사결정자들의 역할이 절실하다. 자신들이 담당하는 업무의 본질을 이해하고,

AI 기술에 대한 이해와 경험이 필요하다. 또한, 데이터 기반의 머신러닝 및 딥러닝 등의 구축 방법을 숙지하고, AI를 활용하여 기획하고 구축하는 능력이 필요하다. 마지막으로, AI를 실제로 적용하여 비즈니스 가치를 창출할 수 있는 능력도 필요하다. 이를 위해서는 현업 의사결정자들의 AI 교육이 시급하다. 코딩 교육이 아니라 AI 교육이 먼저 필요하다. 코딩 없이도 AI를 이해할 수 있다.

SaaS형 AI 활용의 한계 및 대안

끝으로 chatGPT 같은 범용, SaaS 형 AI가 비즈니스 사용에 부적절한 경우 및 이에 대한 대처다. 첫째, 범용 AI는 특정 분야에 사용하기에는 해당 분야 지식이 부족하다. 예를 들어, 삼성전자나 SK하이닉스의 반도체 엔지니어들이 업무에 사용하기 어렵다. 반도체 설계, 제조, 품질 관련 문서가 학습데이터에 포함되어 있지 않았기 때문이다. 다른 아주 특수한 분야도 마찬가지다. chatGPT는 “다양한 분야의 책, 소셜미디어 등을 엄청나게 많이 읽어서, 누가 무엇을 물어보든 막힘없이 그럴듯하게 답하지만, 자신도 자기 말의 진위를 모르는 범용 컴퓨터 SW”이다. 이를 해결하기 위해 해당 분야 지식을 듬뿍 가지고 있는 전문 문서를 모아서 질의어에 추가할 수 있지만, 양이 너무 많기도 하고 기밀 누설이라는 두 번째 문제를 만나게 된다. 삼성전자, SK하이닉스 같은 민간기업은 자신들의 제조 노하우가 OpenAI로 그리고 마이크로소프트로 흘러 들어가는 걸 원치 않는다. 이것은 SaaS에 나타나는 공통점이다. 국정원, 검찰, 심사평가원 같은 국가 기관은 더더욱 곤란하다. 그래서 결국 자신들만의 특정 분야 지식을 학습한 사내에 존재하는 AI가 필요한 것이다. 이러한 AI는 일반적인 대화형 AI보다 높은 수준의 전문성을 가지며, 특정 분야에서 뛰어난 성능을 발휘할 수 있다. 자체 개발을 위해서는 첫째, 사전 학습된 오픈소스 거대 언어 모델을 확보해야 한다. 예를 들어, 빅사이언스의 BLOOM, 메타의 LLaMA, 데이터브릭스의 Dolly 2.0, 모자이크 ML의 MPT-7B 등이다. LLaMA를 제외하고는 모두 상업적 사용도 허용된다. 둘째는 자체 전문 지식을 포함한 학습데이터를 확보하고 이를 이용하여 사전학습 된 언어 모델을 미세조정 하거나 프롬프트 학습을 수행하여, 원하는 태스크에 대한 성능을 최고로 끌어올린다. LangChain은 이러한 자체 개발에 필요한 AI 개발 프레임워크다. 아직은 구축 비용이 너무 높지



만, 이 문제는 기술 발전 속도를 보면 1~2년 내로 어느 정도 해소될 것으로 보인다.

제언

대부분의 기업이 지난 10년간 우리에게 온 빅데이터 기반의 분석 시도 제대로 이해하거나 활용하고 있지 못한 상황에서, 생성 시가 본격적으로 등장했다. 이 분야의 발전 속도는 정말 대단하다. 한두 달이 멀다 하고 새로운 혁신적인 시가 도처에서 출현한다(그림 2. 최근 개발된 AI 리스트). 이 분야 전문가들도 어지러워하고 있지만, 반대로 보면 훨씬 저가에 더 빨리 학습하면서도 성능이 좋아진 시가 한 달이 멀다 하고 개발·출시된다는 뜻이다. 이건 모두에게 큰 기회가 된다. 누가 먼저 이 기회를 깨우치고 적극적으로 도입하느냐가 디지털 변환에서 살아남느냐를 결정할 것이다. 얼마나 빨리 시작해야 하는지에 대한 이견은 있을 수 있으나, 언젠가는 해야 한다는 데는 모두 동의하고 있다.

...	저자소개	↗
<p>조성준 서울대학교 산업공학과 교수는 공공데이터전략위원회 위원장, 한국빅데이터마케팅학회 회장을 역임하였으며, 현재 서울대 빅데이터 AI 센터장과 국가데이터정책위원회 위원을 맡고 있다. 저서로는 『세상을 읽는 새로운 언어, 빅데이터』, 『빅데이터 커리어 가이드북』, 『데이터 천재들은 어떻게 기획하고 분석할까?』가 있다. 서울대 산공과 학사, 미국 메릴랜드대 CS 박사 학위를 받았다.</p>		