

기록관리 이슈페이퍼

ISSUE PAPER vol. 39

국내 기록관리 인공지능 기술 적용 검토



기록관리
이슈페이퍼 vol. 39
I S S U E P A P E R

인공지능 기술의 기록관리분야 적용 방안

방재현 (holybread74@gmail.com /

한국외국어대학교 정보·기록학과 겸임교수, (주)스토리안트 대표)

- I. 들어가며
- II. 인공지능 기술 분류와 기록관리
- III. 인공지능 기술의 기록관리 응용
- IV. 나가며

지속가능한 AI 기술개발 및 기록관리분야 적용을 위한 검토

- AI 기술에 대한 이해 및 기술개발의 과정 -

김용일 공업연구소 (yond@korea.kr / 기록협력과)

- I. 서론
- II. 기술과 데이터의 이해
- III. 기술개발과 도입의 검토과정
- IV. 지능형 기록관리 연구개발 계획
- V. 맺음말

인공지능 기술의 기록관리분야 적용 방안

방재현 (holybread74@gmail.com / 한국외국어대학교 정보·기록학과 겸임교수,
(주)스토리안트 대표)

목 차

- I. 들어가며
- II. 인공지능 기술 분류와 기록관리
- III. 인공지능 기술의 기록관리 응용
- IV. 나가며

『기록관리 이슈페이퍼』는 기록관리 주요 정책과 현안에 대한 열린 논의를 위해 다양한 제언과 연구 결과를 소개하고자 합니다. 따라서 수록된 내용은 국가기록원의 공식적인 입장과는 다를 수 있습니다.

요약

대규모언어모델을 포함한 인공지능 세부분야에 대한 이해를 바탕으로 기록분야에 실무응용 가능한
부분을 검토하기 위해 참고할 내용을 정리한 글임

I. 들어가며

2016년 이른바 ‘알파고 모멘텀’ 후에 사회 전(全)분야에서 인공지능 기술에 대한 기대에 부풀어 있었고 다양한 분야에서 적용하고자 하는 시도가 있었다. 하지만 대중들이 체감하기에는 너무 전문적인 기술분야로 이해하는 것도 어려웠다. 대중이 영화를 통해 알고 있던 그 인공지능의 내용과는 많은 괴리가 있었고 이러한 간극은 쉽게 좁혀지지 않았다. 또한 새로운 기술을 도입해 성과를 내야한다는 조급함은 당시 기술의 이해와는 동떨어진 막연한 기술적 환상에 사로잡혀 있었던 것이 사실이다. 우리 기록분야에서도 이러한 움직임이 있었고 기록관리에 적용하고자 하는 연구가 일부 진행되었다. 하지만 기록관리 업무 적용 가능성에 대한 실증적 연구보다는 이론적이고 가설적인 연구결과만 제시하였고 이것은 실무자들의 기술적 갈등을 해소하기는 부족했다. 인공지능 기술을 선도적으로 도입한 조직과 기관들은 인공지능 기술의 적용 효과를 체감하지 못했다. ‘ChatGPT’가 알려지기 전 인공지능은 제3의 겨울을 맞이할 수 있다는 우려의 목소리가 나타나기 시작했다.

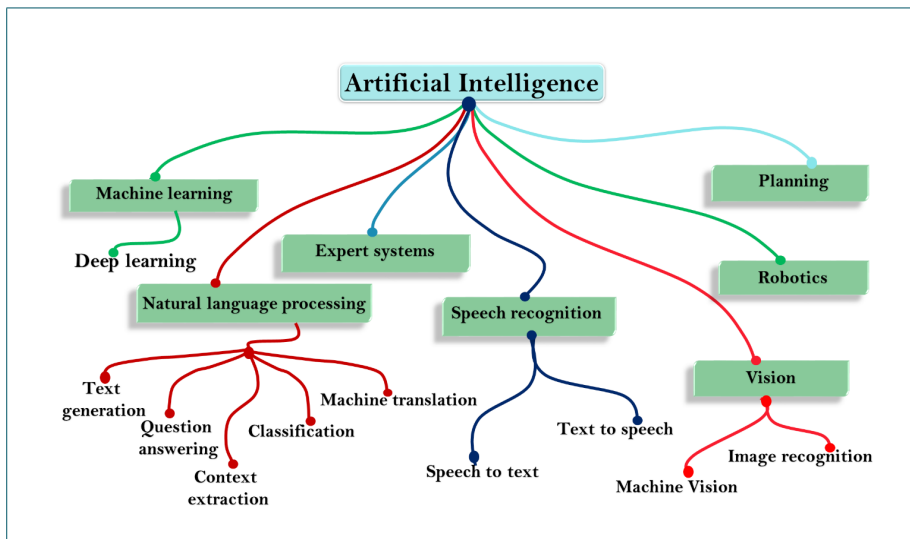
2022년 11월 대화형 인공지능 서비스인 ‘ChatGPT’의 등장으로 인공지능(Artificial Intelligence) 기술은 복잡한 알고리즘을 기반으로 전문가들만 사용할 수 있는 영역에서 누구나 사용할 수 있는 도구로 우리에게 다가왔다. 대화형 인공지능 서비스를 누구나 이용할 수 있었고 이 서비스를 통해서 레포트, 질문, 코딩, 시나리오 작성 등 일반 대중들이 쉽게 도움을 받을 수 있었으며 이것은 인공지능 기술의 실무적 도입이라는 측면에서 또 다른 도화선이 되었다.

이렇듯 점차적으로 고도화되는 인공지능 기술이 기록관리업무에도 어떠한 영향을 줄 수 있을지 고민하고 행동해야 하는 시점인 것은 확실하다.

II. 인공지능 기술 분류와 기록관리

‘ChatGPT’는 대중화된 기술로 우리에게 다가왔고 대형언어모델(Large Language Model)을 기반으로 인간과 대화가 가능한 수준으로 인식되고 있다. 초기에는 부족한 면이 있었지만 보완되어 가면서 점차 발전하고 있는 것도 사실이다. 일반인들은 ‘ChatGPT’에 흥미를 가지고 놀라고 있을 때 이것을 조직이나 업무에 적용하고자 하는 기업이나 사람들은 발빠르게 커뮤니티를 형성하며 발전해 나아갔다. OpenAI에 ‘ChatGPT’가 다국어 지원하기는 하지만 한국어데이터를 기반으로 하는 대형언어모델이 알려지게 되었고 이를 발전시켜 서비스하는 기업들도 급속도로 생겨나기 시작했다. 언어모델은 인공지능 기술분야 중 일부분이며 우리가 잘 알지 못하는 다양한 분야에서도 많은 연구가 이루어지고 있으며 급격한 기술변화가 일어나고 있다.

우리는 인공지능이라는 통칭을 사용하지만 인공지능 기술은 세부분야로 구분되며 각 분야별 연구도 활발하게 진행되고 있다. 기술의 목적별로 인공지능 세부분야에 대한 다양한 기술 체계가 존재하지만 공통적으로 구분되는 분야는 기계학습, 자연어처리, 시각지능, 로보틱스, 전문가시스템으로 구분할 수 있다.



〈그림 1〉 인공지능 기술의 세부분야

출처 : javatpoint.com/subsets-of-ai

기계학습(Machine Learning)은 데이터를 학습하고 해당 학습을 기반으로 예측 또는 결정을 내릴 수 있는 알고리즘 및 모델 개발에 중점을 두는 인공지능의 하위 분야이다. 알파고로 잘 알려진 딥러닝(Deep Learning)은 개선된 학습알고리즘이라고 설명할 수 있다.

자연어처리(Natural Language Processing)는 컴퓨터와 인간 언어 간의 상호작용을 다루는 인공지능의 하위 분야이다. 컴퓨터가 인간의 언어를 이해하고 해석하고 생성할 수 있도록 하는 것이다. 비정형 데이터 형식으로 들어온 음성이나 텍스트 데이터의 입력값을 분석하고, 이를 토대로 결과를 출력하는 것이다. 대용량 문서 처리, 실시간 기계번역, 분석 및 보관, 자동화된 고객 서비스를 위한 챗봇, 텍스트 분류 및 추출, 텍스트 요약 등에 적용되고 있다. 우리에게 알려진 ChatGPT는 자연어처리 기술발전의 산물이라고 할 수 있다.

컴퓨터 비전(Computer Vision)은 컴퓨터를 통해 디지털 이미지, 비디오 및 기타 시각적 입력에서 의미 있는 정보를 추출하고 이러한 정보를 바탕으로 예측, 생성, 분류 등을 수행하는 것이다. 특히 이미지를 분류하고 객체 감지, 객체 추적, 콘텐츠 기반 이미지 검색 등에 적용되고 있다. 우리가 알고 있는 자율 주행 교통 수단에 적용되고 있는 기술이기도 하다.

로보틱스(Robotics)는 자율적으로 다양한 작업을 수행할 수 있는 로봇을 설계, 제작 및 프로그래밍하는 분야이다. 아이작 아시모프의 소설 속에서 ‘로봇을 다루는 기술 분야’를 가리키기 위해 만든 말이었으나 기술의 발전과 함께 학문으로 자리잡게 되었다. 로보틱스의 주요 과제 중 하나는 로봇이 환경을 인식하고 상호 작용할 수 있도록 하는 알고리즘을 개발하는 것이고 최근 인공지능 기술의 발전으로 활발한 연구가 진행되고 있는 분야이기도 하다.

전문가 시스템(Expert Systems)은 특정 분야에서 인간의 의사결정 능력을 모방하도록 설계된 고도의 컴퓨터 프로그램이다. 전문가 시스템은 지식 표현 기술과 추론 엔진의 조합을 사용해 구축된다. 잘 알려진 IBM Watson과 같은 의료 진단 시스템 등이 있다.

이렇듯 인공지능 기술은 다양한 세부 기술분야로 구분된다. 최근 기록학 분야에서도 인공지능 기술 적용 가능성에 대한 연구가 진행되고 실제 적용한 사례가 나타나고 있다. 하지만 대부분의 연구는 적용 가능성에 대한 검토와 이론적 접근방법에 대한 내용이 주류를 이루고 있으며 실제 기록정보를 가지고 적용한 사례들은 오픈된

학습모델을 한정된 데이터셋으로 Fine-tuning¹⁾해 적용하는 사례가 대부분이라고 할 수 있다.

대량의 데이터만 가지고 있으면 기계를 학습시켜 업무에 적용할 수 있을 것 같지만 실제로는 데이터 수집, 데이터 정제, 학습데이터 변환, 파라미터 조정, 테스트 등의 세부 과정들이 존재한다. 실제 데이터를 가공하고 정제해서 기계가 학습할 수 있는 형태로 변환해야 하고 학습모델을 최적화하는데 많은 시간과 자원이 소요되며 해당 도메인(업무분야)에 테스트하고 결과를 분석해 보완하는 과정들을 고려할 때 간단하지 않다.

보통은 새로운 기술을 도입할 때 해당 기술을 조직이나 업무에 어떻게 적용할 것인가 조금하게 고민하게 된다. 하지만 우선되어야 하는 것은 조직이나 업무에서 데이터 복잡도, 형태, 용량, 자원소요 등을 고려해 분석하고 해당 업무를 효율적으로 처리할 수 있는 기술을 적용하는 것이 시행착오를 줄이는 방법이라 할 수 있다.

인공지능 기술을 기록관리에 적용하기 위해서는 기술을 적용하기 위한 업무를 찾는 것이 아닌 현재 기록관리업무에서 데이터를 처리하는데 많은 자원이 소요되는 업무나 과정을 도출해야 한다는 것이다.

예를 들어, 원문검색이 필요할 경우 전자화기록은 일부 광학문자인식(OCR) 처리된 원문도 있지만 대부분은 이미지 PDF 형태로 저장되어 있어 검색이 용이하지 않다. 해당 기록을 검색하고 뷰어를 통해 철/건을 하나씩 열람하는 것은 많은 자원을 소모하게 된다.

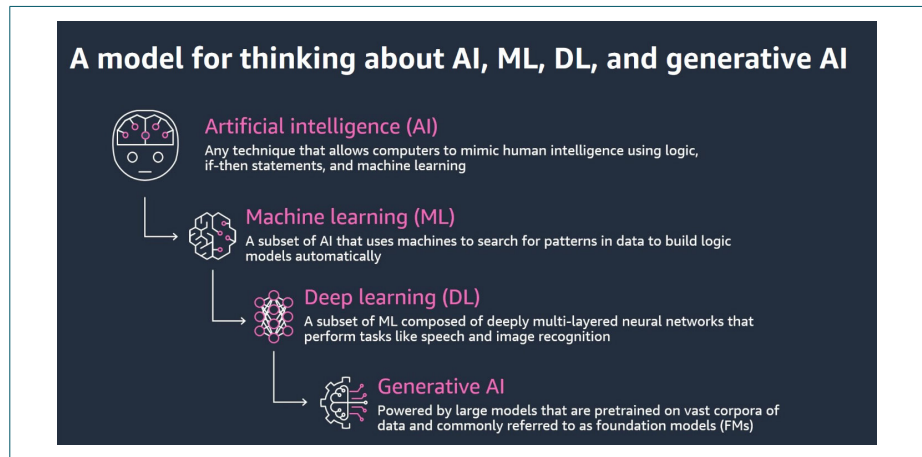
광학문자인식(OCR)기술은 오랫동안 연구되어 온 분야 중 하나이고 모든 문자형태의 패턴과 특징을 인식하기 위해서는 규칙기반 알고리즘으로는 한계를 지니고 있었다. 하지만 딥러닝 알고리즘으로 OCR의 인식률은 향상되었고 현재는 딥러닝 기반 OCR을 통해 문자인식이 필요한 분야에서 활발하게 적용되고 있다.

기록관리시스템에서 전자화된 원문을 검색하기 위해서는 많은 업무시간이 소요되고 이를 효율적으로 처리하기 위한 기술 중 하나가 딥러닝 기반 OCR이라고 할 수 있다. 한국지능정보사회진흥원에서는 AI기술 및 제품·서비스 개발에 필요한 AI 인프라를 지원하고 있으며 학습데이터를 제공하고 있다. OCR관련된 학습데이터셋은 대용량 손글씨 OCR데이터 등 모두 8건의 정제된 학습데이터가 제공되고 있다.

1) Fine-tuning은 사전 훈련된 학습모델의 가중치가 새로운 데이터에 대해 훈련되는 전이학습에 대한 접근 방식으로 이미 학습된 모델을 새로운 데이터셋에 대해 재학습하는 것

III. 인공지능 기술의 기록관리 응용

기록관리분야에서 많은 부분을 차지하는 데이터 형태는 텍스트이다. 텍스트 데이터를 기반으로 가장 주목받는 기술은 아마도 생성형 AI일 것이다. 또한, 2023년에 인공지능 분야에서 가장 독보적인 세부기술분야는 생성형 AI이다. 생성형 AI는 아래 그림과 같이 Deep Learning의 하위 기술영역이다.



〈그림 2〉 인공지능, 기계학습, 딥러닝, 생성적 AI의 구성

출처 : <https://aws.amazon.com>

특히 생성형 AI 중에서 ChatGPT가 촉발시킨 대화형 AI에 적용된 대형언어모델은 정제된 텍스트 데이터를 대량으로 보유하고 있는 기록분야에서는 적극적으로 활용방안을 마련하고 연구해야 하는 분야이기도 하다.

인공지능 기술이 기록관리에서 활용되기 위해서는 실제 업무담당자의 업무를 얼마만큼 경감시켜줄 수 있느냐가 관건이라고 할 수 있다. 물론 최종 의사결정을 인공지능 기술에 맡기는 것은 당분간 어려울 것으로 판단된다. 그렇다면 우리는 이러한 인공지능 기술을 어떻게 활용할 수 있을 것인지 고민해야 할 것이다.

예를 들어 기계가 가독할 수 있는 데이터만 있으면 이를 신속하게 분석하고 업무담당자의 의사결정에 유의미한 결과를 제공해 업무부하를 경감시킬 수 있을 것이다. 신속한 기록물의 검색과 내용파악을 위해서 요약정보가 필요하다. 하지만 메타데이터 항목에 요약정보가 입력되어 있는 기록은 그렇게 많지 않다. 100페이지 분량의 원문내용을 요약하는 것은 기록관리업무에 부하를 주는 것이 사실이다. 도메인에 특화된 대형언어모델을 통해 1차적으로 원문내용요약을 생성할 경우 검색이나 주요 키워드

추출 등 기록관리 및 서비스를 위한 새로운 접근방식을 제공할 것이다.

또는 정보공개가 요청된 기록의 경우 해당 원문의 비공개정보 식별은 업무담당자의 업무부하로 작용하고 있다. 각 기관이나 조직이 보유하고 있는 비공개대상정보나 가이드를 기계가 학습해 1차적으로 공개요청에 대한 식별을 업무담당자를 위해 지원하게 될 경우 신속한 대국민서비스를 실현할 수 있을 것이다.

기록관리영역에서 인공지능 기술을 적용할 수 있는 업무영역은 다양하다. 기록관리 영역에서 담당자의 업무부하가 가장 많고 검토해야 할 데이터가 많은 부분을 우선으로 정리한 표는 아래와 같다.

〈표 1〉 지능형 기능 영역에 따른 기계학습 기술 영역

기능 영역	기술(Technology) 영역
업무유형분류, 보유일정, 민감정보 분류, 공개구분, 내용분류, 분류체계	자동 분류 (Auto Classification)
문맥분석, 내용요약, 기록정보 검색, 기록물 기술(Description)	문서 요약 (Text Summarization)
이미지 태깅, 이미지 내용 기술(Description), 이미지 검색	이미지 자동 태깅 (Image Auto-Tagging)
기록콘텐츠 생산, 기록정보서비스	콘텐츠 생성(Contents Creation)
검색, 시청각 기록 분류	내용기반 이미지 검색 (Content-based Image Retrieval)
업무분석, 분류체계수립	군집화 (Clustering)

대형언어모델을 기반으로 문서의 문맥을 분석하고 내용을 요약하는 것은 현재는 보편화된 기술이다. 이미 보편화되어 응용되고 있는 기술을 기록관리영역에서 적용가능성을 검토하는데 시간을 소요하기 보다는 우선 적용한 후 결과를 도출하는 것이 좀 더 현실적일 것이다. 인공지능 기술을 적용하는 것이 실제 도메인에 존재하는 정제된 데이터셋을 이용하고 추론의 정확도 향상을 위한 반복적 학습과 같은 기본적인 단계 이외에도 적용과정에서 발생하는 여러 가지 예외사항을 고려할 때 이론적이고 가설적인 접근방식은 적합하지 않다. 즉, 적용가능성 보다는 적합한 언어모델을 도메인에 맞춰 Fine-tuning하고 해당 모델을 통해 업무효율성, 정확도, 적용요구조건을 도출해 실무에 적용하는 단계를 단축하는 것이 좀 더 효율적인 방안이라 생각된다. 우선은 기록관리업무에서 인적자원의 한계, 물리적 소요시간, 검토할 데이터의 양을 고려해 업무부하도를 측정하고 이를 해결하기 위한 도구로

인공지능 기술을 도입하는 것이 바람직하다. 인공지능 기술이 인간의 사고영역을 언제 침범할지 모르지만 대규모 데이터의 분석이나 처리, 이를 기반으로 하는 예측의 영역은 이미 인간의 수준과 비교조차 할 수 없다.

IV. 나가며

인공지능 기술이 우리에게 알려진 것은 최근이지만 인공지능 기술이 연구된 것은 컴퓨터 등장과 함께 진행되어 왔다. 인공지능 기술이 1950년대부터 두 차례의 침체기를 겪었던 가장 큰 이유 중 하나는 인간의 과도한 기대 때문이었다.

인공지능 기술에 대한 재전성기가 지속되고 있는 지금, 공상과학영화에서 나오는 인공지능 기술을 기대하는 이에게는 실망스럽겠지만 대규모 데이터를 생산하고 관리하는 기업이나 조직에서는 이보다 더 유용한 도구는 없을 것이다.

인공지능 기술을 업무를 대신하는 기술로 받아들이기보다 업무를 효율적으로 처리하기 위한 하나의 도구로 받아들여야 할 것이다. 지금이라도 기록분야에서는 인공지능 기술을 적극적으로 도입하기 위한 실증적 연구가 활발히 진행되어야 하고 막연한 기대를 가지고 접근하는 것보다 인공지능 기술의 장점과 한계를 명확히 구분하고 세부기술에 대한 이해를 바탕으로 응용연구를 진행해야 할 시점이다.

◆ 필자소개

현재 한국외국어대학교 정보·기록학과 겸임교수로 디지털 아카이브 시스템, 기록콘텐츠, 인공지능을 강의하고 있으며, 지능형 아카이브 시스템을 연구/개발하는 (주)스토리안트 대표로 재직하고 있다. 컴퓨터공학(인공지능) 전공으로 석사, 지능형 아카이브 시스템으로 기록학박사 학위를 받았다.

지속가능한 AI 기술개발 및 기록관리분야 적용을 위한 검토

- AI 기술에 대한 이해 및 기술개발의 과정 -

김용일 공업연구사(yond@korea.kr / 기록협력과)

목 차

I. 서론

II. 기술과 데이터의 이해

1. AI 기술 이해의 필요성
2. 딥러닝의 유행과 경쟁
3. 사전학습 모델
4. 빅데이터와 인공지능
5. 데이터의 목적

III. 기술개발과 도입의 검토과정

1. AI 문제의 정의
2. 기술사이클의 참조
3. 표준 · 지침의 활용
4. 지속가능한 인공지능 개발

IV. 지능형 기록관리 연구개발 계획

V. 맺음말

『기록관리 이슈페이퍼』는 기록관리 주요 정책과 현안에 대한 열린 논의를 위해 다양한 제언과 연구 결과를 소개하고자 합니다. 따라서 수록된 내용은 국가기록원의 공식적인 입장과는 다를 수 있습니다.

요약

기록관리 분야의 인공지능기술 적용에 앞서 필요한 기록관리 구성원간의 AI 기술에 대한 이해 및 공감대의 구축과 기획-연구개발-기록관리 적용 과정에 유용한 다양한 방법론과 프로세스에 대한 내용을 정리한 글임

I. 서론

4차 산업혁명이라는 단어가 2016년부터 유행하고, 알파고가 사회에 충격을 가져다준 이후로 IT 기술은 거의 7년째 트렌디한 주제로 자리매김하고 있다. 하지만 전문화된 영역에서 AI(Artificial Intelligence)의 성공적 도입과 활용사례는 많은 편은 아니다. 그 이유는 선진화된 기술이 생산성을 높이고, 새로운 영역의 발견을 가능하게 하는 것은 사실이나 어떤 기술을 사용해야하고, 그 기술을 사용하기 위해 어떤 준비가 필요한지에 대해서는 기술개발 만큼의 고민이 필요하기 때문이다. 국가기록원에서는 4차산업 기술과 관련된 사전조사 및 내부검토를 통해 중장기 계획을 세우고 기록관리 영역과 관련된 지능형 기술 개발을 추진하고 있다.

본 글에서는 AI 기술 동향에 대한 간략한 소개와 더불어 기술의 개발·도입을 검토하는 과정, 그리고 국가기록원의 계획을 소개한다. 이 글을 통해 성공적인 지능형기술 연구개발 및 기록관리 적용을 위한 계획과 관련된 내용을 공유하고 기록관리와 관련된 구성원의 공감대를 갖고자 한다.

II. 기술과 데이터의 이해

우리는 인공지능에 관한 얘기를 할 때 각자 인공지능의 범위와 역할, 기대수준에 대해 의견 차이가 존재한다. 인공지능이 가까운 미래에 많은 것을 해결할 수 있다는 긍정적인 의견부터 모험적인 투자가 필요하다는 의견, 도덕적 관점의 신중론 등 다양한 논지의 의견이 등장할 때도 많다.

현재 인공지능 기술의 수준과 그 효과에 대해 공감대를 가지는 것이 중요하다.

우리는 인공지능이라는 주제를 다루면서 구체적 실현 방법이 없이 단순한 거시적인 계획만 존재하거나, 기술적인 수식어와 용어의 홍수로 인해 본질을 벗어나 미시적인 부분에만 매몰되는 일도 흔하다. 기술과 그 개발과정에 대한 이해를 토대로 알려진 실패를 반복하지 않기 위해 기술의 동향을 파악하고, 현실적으로 가장 효과적인 방법을 적용하기 위한 방법을 찾아야 한다.

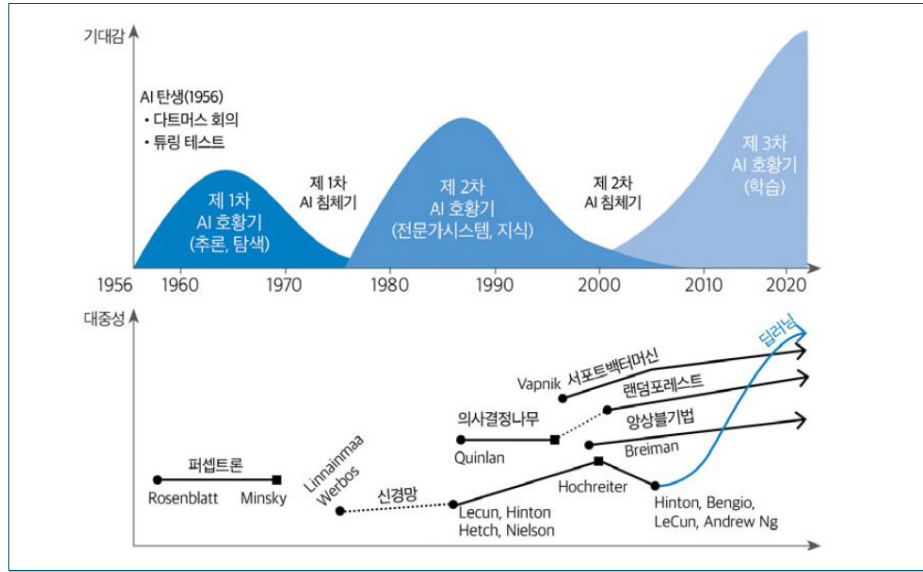
1. AI 기술 이해의 필요성

많은 강연과 교육에서 인공지능을 소개하며 지능의 수준, 문제해결 범위에 따라 약인공지능(Narrow AI, Weak AI), 강인공지능(General/Strong AI), 초지능(Artificial Superintelligence) 3가지 유형으로 분류한다. 현재 실현가능한 인공지능은 약인공지능 수준으로 특정 기능(자율주행-운전, 알파고-바둑 등)만이 수행가능한 인공지능이다. 사실상 범용인공지능과 더 나아가 초지능은 공상과학과 철학의 영역으로 현실 문제에서는 아직 '강인공지능이 온다면'이라는 가정조차 무의미한 단계이다. 현실 문제를 해결하기 위해서는 이러한 추상적인 개념에 대해 논하기보다는 실제적인 기술에 초점을 맞추는 것이 바람직하다고 본다.

2016년 구글 딥마인드의 알파고 등장 이후, 두차례의 암흑기를 거쳤던 AI기술이 재조명되면서 4차 산업혁명이라는 단어와 함께 기술에 대한 폭발적 관심과 투자가 약 7년째 이어져오고 있다. 2015년 이전까지 인공지능 기술은 학술적이고 전문적인 영역으로 수학, 통계학, 언어학 등 중심의 학문이었다. 알파고라는 이벤트를 통해 대중적 관심 또한 폭증한 이후, 전문가시스템이라는 영역에서 개개인의 4차산업혁명에 대한 소양, 데이터 분석역량 또한 강조되고, 기술에 대한 이해까지 요구되는 것으로 확장되었다. 그렇다면 기술을 어느 정도까지 이해해야 하는가 의문이 든다. 모두가 기술개발을 수행할 필요는 없지만, 인공지능의 기본 원리와 흐름을 이해하고, 역할과 한계를 파악한다면 현실의 문제를 해결하기 용이할 것이다.

2. 딥러닝의 유행과 경쟁

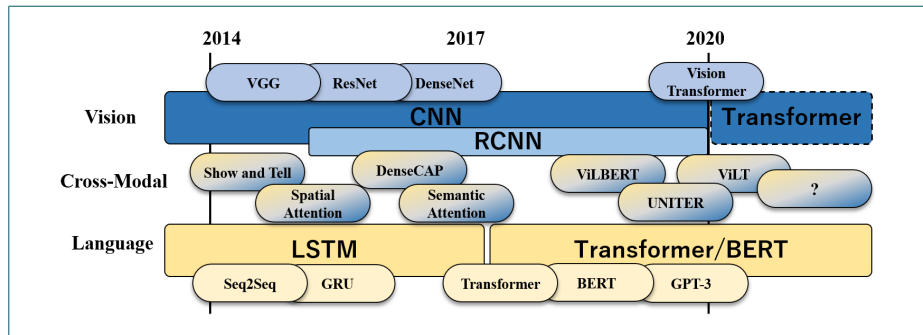
기술의 발전흐름을 거슬러보면 인터넷 보급 이후 IT 기술의 패러다임은 초창기에는 데이터 검색이 중요했으나 데이터가 충분히 축적되고난 이후에는 데이터 분석에 대한 수요가 증가하였다. 처음에는 확률·통계 기반의 기계학습 알고리즘을 이용하여 전문적인 데이터과학자만 기계학습이라는 주제를 다루었지만, 현재는 심층인공신경망(Deep Neural Network)인 딥러닝이 모든 영역을 압도하고 있고, 오픈소스 프레임워크가 고도화되면서 그 사용 또한 간편해지게 되어 많은 소프트웨어 개발자들이 인공지능 소프트웨어를 개발하고 있다.



〈그림 1〉 인공지능에 대한 관심 및 기술 대중화의 흐름

출처 : 빅데이터 · 인공지능 산업 진흥을 위한 데이터 과학의 발전 전략 연구(한국과학기술한림원, 2019)

언어모델, 강화학습, 생성모델 등 현재 접할 수 있는 모든 인공지능 모델은 인공지능경망 기반이다. 인공지능경망은 두뇌의 뉴런을 모방한 모델로 과거에는 계산량이 많아 사용하기 어려웠지만 GPU(그래픽처리장치)를 이용한 병렬계산 방법이 활성화된 이후 모든 인공지능 문제에서 사용되고 있다.



〈그림 2〉 인공지능경망 기술의 발전 흐름

출처 : Perspectives and Prospects on Transformer Architecture for Cross-Modal Tasks (Andrew Shin)

2016년도까지는 인공지능경망을 토대로 순환신경망(RNN, 텍스트 AI에 주로 사용), 합성곱신경망(CNN, 이미지AI에 주로 사용) 구조를 기본으로 그 응용형태에 따라 GRU, RCNN, Attention-RNN 등 인공지능경망의 구조를 발전시켜 왔다.

2017년 구글에서 ‘Attention is all you need’라는 유명한 논문을 발표한 뒤로는 이 논문에서 제시된 트랜스포머(Transformer) 모델이 거의 모든 분야에 걸쳐 인공지능경망에 대한 성능이 우월하였고, 현재 독보적으로 많이 사용되고 있다.

3. 사전학습 모델

인공지능경망의 학습에는 데이터가 많이 필요하고, 데이터의 균형 또한 중요하다. 그리고 학습에 투입되어야하는 비용이 매우 크기 때문에 대기업 등에서 미리 학습된 인공지능경망을 제작하여 배포하였는데 이를 사전학습(Pre-trained) 모델이라고 한다. 그리고 이 사전학습 모델을 사용자가 가진 고유의 데이터에 적용하는 것을 전이학습(Transfer Learning)이라고 한다. 현재 대개의 AI 적용은 사전학습모델을 활용한 전이학습에 초점이 맞춰져 있다. 과거에는 사용자가 가진 많은 양의 데이터로 전체 AI 모델을 미세조정(Fine-tuning)하는 방법이 쓰였으나, 기계학습 모델이 대형화됨에 따라 미세조정에 필요한 데이터를 충분히 구비하는 것이 어려워져 적은 데이터로 추가적인 학습을 하는 퓨샷(Few-shot)러닝 기법이 널리 쓰이고 있다.

Chat-GPT의 GPT 3.5 역시 OpenAI社에서 제작하여 배포한 사전학습 대형 언어모델이다. GPT 3 버전까지는 소스코드를 모두 공개했지만 현재 최신 버전은 공개하지 않고 있다. 사용자의 데이터로 모델을 최적화시키기 위해서는 GPT에서 제공하는 API를 이용해 퓨샷러닝 등의 작업을 수행해야한다. 물론 학습모델을 전체 공개한다고 하더라도 거대한 AI 언어모델을 구동할 전산자원과 언어모델에 학습시킬 데이터를 구비하는 것이 대기업이 아니고서는 불가능하다. 현재까지는 대부분의 인공지능 모델이 오픈소스로 공개되었지만, GPT 3 이후부터는 공개가 잘 되지 않는 추세로 대기업의 대형 언어모델의 사용에 대한 과금이라는 상업구조로 개편될지 모른다는 전망도 있다.

GPT와 같은 대형언어모델(LLM)은 높은 구동비용으로 온-프레미스(자체구축)가 불가능하고 원격 API 방식의 학습과 사용을 추구하다보니 외국 IT기업의 의존과 데이터 유출에 대한 우려가 큰 상황이다. 이로 인해 공공분야에서는 LLM을 경량화한 sLLM(small-LLM)의 개발 및 도입을 통해 의존도 및 보안 문제에 대응하려고 하고 있다. 국가기록원에서도 초거대 AI 학습을 위한 말뭉치 데이터 구축사업에서 구축되는 데이터를 기반으로 향후 연구개발사업을 통해 sLLM에 대한 연구 및 그에 대한 활용방안을 수립할 계획이다. sLLM의 기본적인 작동 원리는 LLM과 유사하지만

파라미터(두뇌의 뉴런에 해당)의 수를 축소하고, 학습데이터의 범위를 범용이 아닌 특정 영역으로 한정된 모델이다. 어느 정도 성능적인 측면에서 타협이 있을 수 있지만, 비용과 실익을 따졌을 때 합리적인 방안이라고 볼 수 있다. sLLM의 자체적인 구축은 대체로 데이터의 범위가 관련분야로 제한되므로 학습데이터의 오염(학습데이터와 테스트데이터가 섞이거나 매우 유사한 특성을 갖는 것 등)과 과적합(Overfitting, 특정데이터에만 예측이 정확한 경우)이 일어날 수 있다. 이러한 경우 개발단계에서 드러나지 않은 문제가 실제 운용시 두드러질 수 있으므로 성능평가 및 검증에 대한 강화가 필요하다.

현재는 인공지능망 모델과 학습데이터가 대형화됨에 따라 새로운 기술을 개발하고 앞서나가기에는 그 장벽이 너무나 커진 모양새이고 현재는 출시된 사전학습모델의 각 분야 적용을 위한 연구가 활발하게 진행 중이다. 그렇기에 현 시점에서는 기술의 활용에 초점을 맞추고 최대한의 효과를 이끌어내는 것이 목표라고 할 수 있겠다.

4. 빅데이터와 인공지능

빅데이터와 관련된 기술은 2002년 디지털로 저장된 정보의 양이 아날로그를 추월하고, 2006년 대표적인 빅데이터 분산병렬처리 프레임워크인 아파치 오픈소스 프로젝트 하둡의 0.1 버전이 발표되면서 본격적으로 시작했다. 이후 사물인터넷(IoT), 스마트폰의 보급, UCC 등과 함께 데이터의 양이 폭발적으로 늘어 현재는 데이터에서 정보와 지식을 추출하는 관리·분석 기술로서 4차산업혁명 기술에서 아주 중요한 부분을 차지하고 있다.

〈표 1〉 빅데이터 기술의 분류

출처 : ICT R&D 기술로드맵 2023, 정보통신기획평가원, 2018

	소분류	세분류	요소기술
빅데이터	빅데이터 처리·유통	빅데이터 수집·유통기술	빅데이터 수집/정제/융합/가공, 데이터 품질관리, 실시간 ETL/ELT, 데이터 생성/증강, 비식별화 및 필터링, 데이터 마켓, 데이터 레이크, 데이터 프리퍼 레이션, 마스터 데이터 관리
		빅데이터 저장·처리·관리 기술	실시간 스트림 처리, 데이터 처리 및 관리, 데이터 라이프 사이클, 가상 데이터맵 관리, 데이터 처리 프레임 워크
	빅데이터 분석·활용	빅데이터 분석·예측 기술	심층 분석, 실시간 분석, 그래프 분석, 예지분석, 예측분석, 지시적 분석, 인지 분석, 시공간 복합분석, 분석 모델링, 시뮬레이션, 데이터 기계학습/딥러닝, 이상치 검출
		빅데이터 활용·시각화 기술	빅데이터 서비스, 빅데이터 응용, 사회변화 예측, 데이터 시각화, 분석 시각화, 대쉬보드

빅데이터의 대표적 처리기술로는 분산 병렬 데이터 처리기술 표준인 맵리듀스가 있으며 분산처리를 위해 Hadoop(빅데이터 저장시스템), NoSQL(비정형데이터베이스)를 기본으로 여러가지 데이터 관리·분석용 기술들이 모여 하둡 생태계(Hadoop Ecosystem)를 이루는 것을 표준으로 하고 있다.

그리고 여러 기업에서 하둡을 기반으로한 빅데이터 플랫폼을 개발하여 선보이고 있으며, 문제는 서로 상이한 플랫폼으로 인해 각 플랫폼간의 공유와 통합은 다소 어려움을 겪고 있다. 이 상이한 플랫폼 간 데이터 교환과 공유를 위해 빅데이터 시스템 연동에 관한 국내 정보통신표준 또한 2017년 개발된 바 있다.

5. 데이터의 목적

인공지능을 이야기할 때 항상 빠지지 않는 것이 데이터의 중요성이다. 빅데이터라는 이름으로 학습데이터 활용, 새로운 인사이트 발견을 위해 막연하다고 할 수 있을 정도로 그 중요성이 강조되고 있다. 인공지능 기술의 성능은 결국 데이터가 좌우하기에 충분한 양의 양질의 데이터를 확보하는 것이 가장 중요하다.

전통적인 데이터베이스 관리와 빅데이터의 차이는 단지 데이터 구조화 여부(비정형데이터 유무)와 처리방법(SQL과 NoSQL)에 있다. 이외에 데이터 품질관리, 거버넌스, 데이터 통합, 검색 등은 빅데이터 기반 시스템에서도 여전히 중요한 영역이며 여전히 대부분의 업무시스템에서는 관계형데이터베이스를 필요로 하고 있다. 기존 시스템(Legacy)이 최적화되고 데이터가 무결하게 유지되고 있을 때 비로소 빅데이터 시스템을 구축할 수 있고 의미있는 운용이 가능하다.

또한 빅데이터 분석 관점에서의 데이터처리와 인공지능 관점에서의 데이터는 약간 다를 수 있다. 빅데이터 분석 관점에서는 전체 보유데이터를 대상으로 데이터 수집·저장·관리·전처리·적재의 과정과 그 이후의 데이터 분석과 탐색, 시각화 단계로 나누어진다. 반면, 인공지능 관점에서의 데이터처리는 학습시킬 기계학습 모델에 최우선적으로 초점을 맞추어 오로지 기계학습에 필요한 데이터 수집·데이터 가공 과정에 한정된다. 그러므로 데이터 처리 프로세스는 전체 데이터 관리와 인공지능 학습데이터의 공통적 측면을 찾아 종합적인 측면에서 검토되어야 한다.

어떤 인공지능 기술을 도입하려고 할 때 데이터 준비에 대해 많은 비용이 수반된다는 것은 알고 있지만 얼마만큼의 데이터와 어떤 형식의 데이터가 필요한지, 또한 그 효과가 어떠한지에 대해서는 정답이 없다. 충분한 데이터의 양은 목표하는 인공지능 모델과 데이터의 특성에 따라 달라지게 되어 고비용의 데이터구축 이전에 이에 대한 면밀한 검토가 필요하다.

앞에서 서술했듯 인공지능 관점에서의 데이터는 학습에 초점이 맞추어져 있으므로 학습시킬 인공지능 모델이 미리 구체적으로 정의되어 있지 않으면 의미가 없게 된다. 또한 성공적인 인공지능 기술 도입에 필요한 데이터에는 몇 가지 조건이 있다. 가장 중요한 것은 데이터의 품질로 데이터가 정확하고, 신뢰성이 있으며, 오류에서 자유롭고 일관성이 있어야 한다. 결론적으로 인공지능 도입에서는 문제의 정의가 가장 중요하다고 볼 수 있으며 미리 인공지능을 이용하여 어떠한 일을 해결할 것인지 사전에 구체적으로 계획되는 것이 필수이다.

III. 기술개발과 도입의 검토과정

생산성 향상과 새로운 가치 창출을 위해 인공지능 도입이 필요하고, 이를 위해선 최적의 기술 채택과 데이터의 구비가 중요하다는 것을 모두가 알고 있다. 실제 도입에 앞서 미리 그 과정을 설계하고 효과를 예측하지 않고 시스템 개발에 착수를 하게 되면 그 시행착오에 대한 리스크와 비용이 커질 수 있다. 이를 위해 기술을 채택하고 인공지능 도입을 통한 문제해결 프로세스 설계에 대한 보편적인 방안들이 고민되고 있고, 표준이나 가이드가 제공되고 있다.

1. 문제의 정의(Problem Definition)

알파고, 챗GPT 등 인공지능 기술이 이슈가 될 때 각 분야에서는 이에 대한 적용 방안을 찾게 된다. 최신 인공지능 기술 검토를 통해 당면한 여러 가지 문제를 통합적으로 해결할 수 있는 솔루션을 찾지만 안타깝게도 현재의 인공지능 기술은 범용적인 문제를 한번에 해결하지는 못한다. 도입을 통한 효과를 보기 위해서는 무엇보다 문제를 구체적으로 정의하는 것이 중요하다. 인공지능을 위한 데이터 구축에 있어 목적을 정의하고 진행하는 것은 인공지능 기술 분야에 대한 적합한 데이터 수집을 통해 신뢰성 있는 결과를 확보하는데 매우 중요한 요소라고 인공지능 학습용데이터 구축안내서(2023, NIA)에서도 데이터 구축 공정의 최상단에서 언급하고 있다.

AI의 선도기업인 구글에서는 여러 가지 방법론들과 적용사례를 통한 문제 정의와 AI 시스템 설계에 대한 도움을 주는 PAIR(People+AI Research) 가이드라인을 제공하고 있다. 사용자 요구사항 정의, 데이터 수집·평가, 설명가능성, 피드백, 오류 등 AI 개발에서의 6가지 단계적 고려사항으로 구성된 가이드북과 함께 각 단계에서 실질적으로 설계에 활용할 수 있는 워크시트가 제공된다. 첫 번째 사용자 요구사항 정의 편에서는 ① AI 기술의 강점과 사용자 요구사항의 교집합 찾기 ② 자동화(Automation)와 증강(Augmentation)의 선택 ③ 성공/실패 리워드의 설계와 평가로 과정을 나누고 있는데 AI 기술의 강점을 가질 수 있는 경우와 AI가 잘 동작하지 않는 경우를 구체적으로 기술하고 있다.

〈표 2〉 AI의 적용 여부 검토

출처 : 구글 People-AI Guidebook

AI probably better	AI probably not better
□ The core experience requires recommending different content to different users. □ The core experience requires prediction of future events. □ Personalization will improve the user experience.	□ The most valuable part of the core experience is its predictability regardless of context or additional user input. □ The cost of errors is very high and outweighs the benefits of a small increase in success rate.

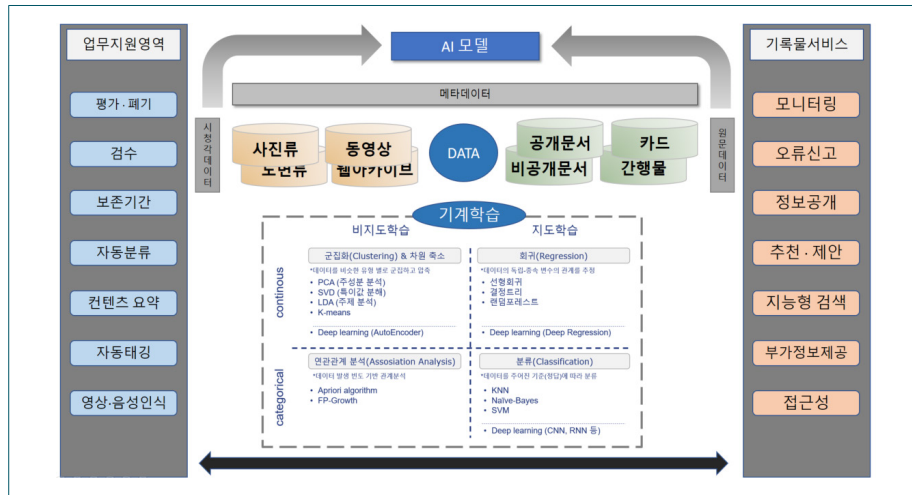
데이터 수집 · 평가 단계에서는 고품질의 데이터 수집계획, 사용자 요구-데이터 요구로의 전환 등 학습데이터로부터 어떻게 사용자 요구에 맞는 인공지능을 구현할 것인지에 대한 단계적 활동에 대한 정보를 제공한다.

〈표 3〉 AI 문제의 정의

출처 : 구글 People-AI Guidebook

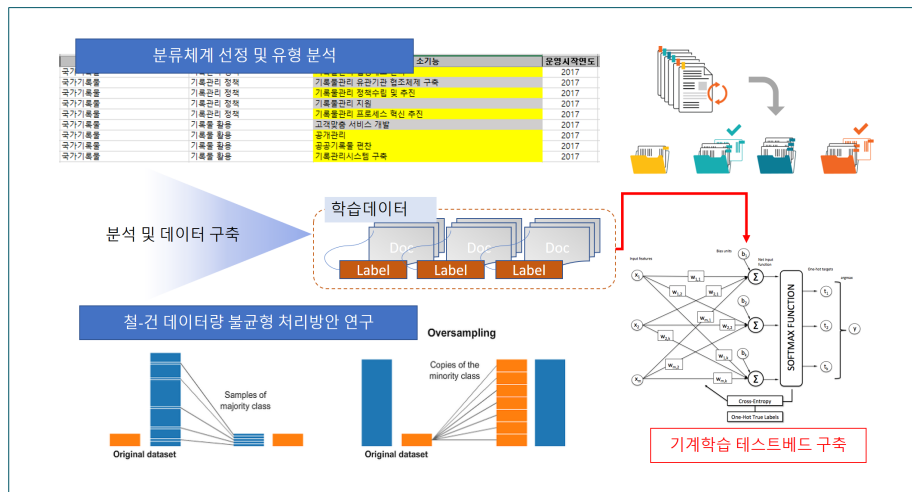
User needs & data needs	
Users	<i>Home chefs</i>
User action (core value prop)	<i>Cook a new dish using the recipe based on recommendation</i>
ML system output	<i>Recommendations for new recipes</i>
ML system learning	<i>Patterns of behavior around choosing recipe recommendations</i>

이처럼 구글 PAIR 가이드에서는 AI 도입에 있어 순차적으로 고려해야할 필수 사항들에 대해 간단하지만 유용한 정보를 제공하고 있다. PAIR를 운영하는 구글 내 조직에서는 기계학습을 보다 설명가능하고 신뢰성있고 유용하게 만들 수 있는 HCI(인간-컴퓨터 상호작용), 디자인 등 AI 서비스가 사용자에게 성공적인 경험이 되기 위한 연구 또한 진행하고 있어 기획 단계에서 참고하면 도움이 될 것이다.



〈그림 3〉 기록관리 분야에서의 AI 기술 연구개발 검토

국가기록원에서는 2017년에 지능형 전자기록관리 기술연구 개발 연구동향 분석 및 수요조사를 통해 4차 산업혁명의 핵심 기술을 활용한 국내외 전자기록관리 자동화 기술의 연구·개발 동향 분석 및 수준 진단을 수행한 바 있다. 이 연구 결과를 토대로 전자기록에 적용 가능한 지능화 기술 유형을 분류하고, 연구개발 동향 분석과 인수, 정리, 분류, 검색, 활용 등 기록관리 분야별 자동화 기술개발계획을 수립하였다.

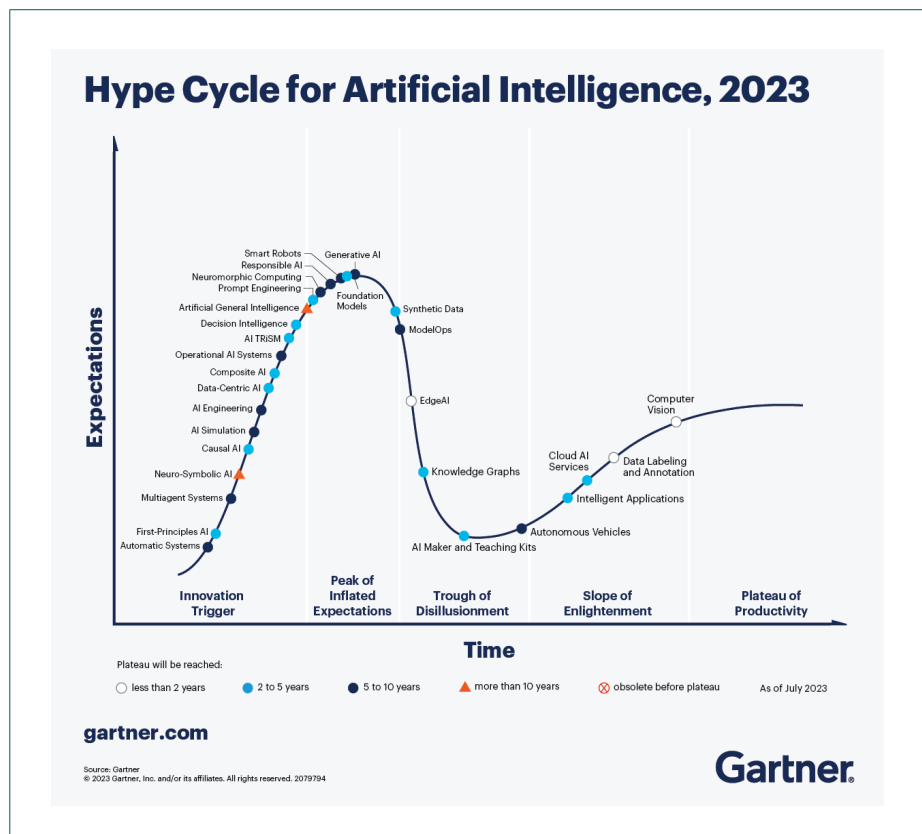


<그림 4> 기록물 분류체계 및 분류자동화 연구

2. 기술사이클의 참조

최근 Chat-GPT가 주목받으면서 생성형 AI가 기술도입의 트렌드가 되고 있다. 하지만 이슈가 되고 있는 최신기술을 우선 고려하고 도입하는 것이 항상 성공을 보장하지는 않으며, 오히려 기술의 불완전성과 개인정보 침해·정보유출 등에 대한 우려가 있을 수 있다. 실험적 단계가 아닌 실제 기술을 도입하는 시점에서는 성숙단계에 있는 효과성이 담보된 기술을 도입해야 한다.

매년 발간되는 미국의 IT 자문회사인 가트너(Gartner)의 하이프 사이클 보고서는 신기술 도입을 위한 의사결정에 참고할만한 정보를 제공한다. 기술들이 혁신적 발단, 과장된 기대를 거쳐 환멸, 계몽, 안정의 사이클을 거친다는 경험적 사실을 토대로 최신 기술의 트렌드 분석하였다. 개별 기술의 특성에 대한 반영이 미흡하다는 비판이 있지만 신기술 도입에 앞서 기술의 성숙도와 효과에 대한 예측정보를 제공한다.

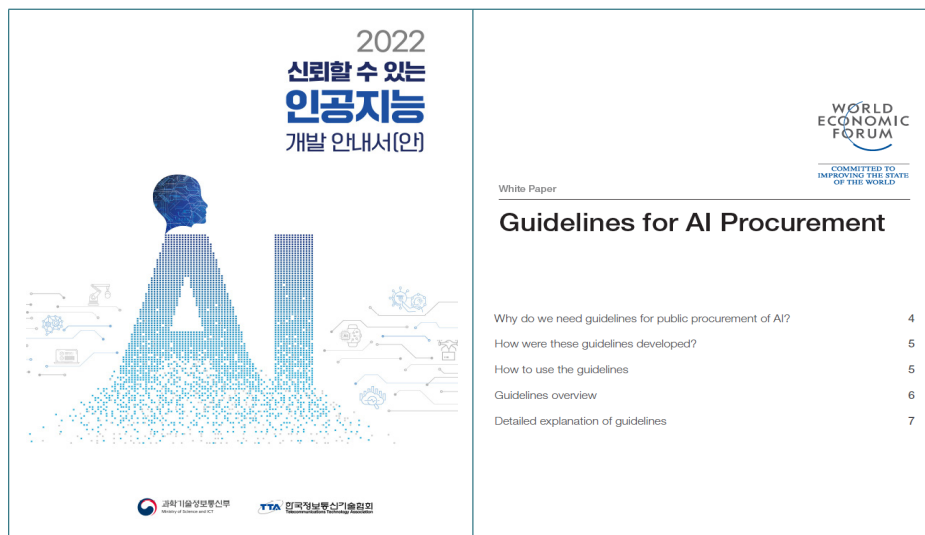


〈그림 5〉 가트너 AI-하이프사이클, 2023

이 보고서에서는 2023년 현재 생성형AI와 파운데이션 모델(대형언어모델)에 대한 기대치가 최고조임을 알 수 있다. 반면 자율주행 자동차에 대해서는 기술한계, 관련법, 안전성 등 실제 적용에 대한 논란이 많고, 아직 해결해야할 과제가 많은 것이 드러나 시대를 앞서가던 기대 자체는 줄어든 상황이다. 이러한 자료를 활용하는 것이 해당 기술에 대한 실제 효과와 파급력을 정확히 예측할 수는 없지만 해당 기술을 도입하는 것이 적극적이고 실험적인 도입인지 실제 서비스를 위한 목적인지 구분에 도움을 줄 수 있다.

3. 표준 · 지침의 활용

효율적인 도입과 높은 효과를 위해서는 인공지능 구축 · 도입에서 단계별로 수행해야하는 주요 활동과 그 결과물에 대한 표준 및 가이드를 활용할 필요가 있다. 신뢰할 수 있는 인공지능 개발과 효율적인 데이터 구축, 공공분야에서의 인공지능 활성화 등에 대한 정책적 연구가 수행되고 사업수행을 위한 가이드 또한 지속적으로 제작 · 배포되고 있다.



〈그림 6〉

신뢰할 수 있는 인공지능 개발 안내서(TTA)

〈그림 7〉

공공부문 AI 조달가이드라인(세계경제포럼)

인공지능 기술의 산업 · 사회적 영향력이 증가하고, 이를 응용한 서비스 개발의 수요가 증가하면서, 관련 기술 표준화에 대한 필요성이 제시되었다. 이에 따라 인공지능 표준화가 2017년부터 국제적인 시작 추세를 보였으며, 현재 인공지능의 용어 및 개념 정의를 위한 표준부터 응용기술 표준까지 다양한 목적으로 표준 개발이 이루어지고

있다. 공적표준화 기구와 사실표준화 기구의 표준 개발 목적이 서로 차이가 있기 때문에, 각 기구에서 현재 개발 중인 인공지능 표준 문서들의 개발 방향은 각기 다른 특징을 가진다. 일반적으로 공적표준화 기구에서는 사회의 다양한 분야에서 표준기술을 반복적으로 활용할 수 있도록 참조 표준을 개발하고 있으며, 따라서 인공지능 기술과 관련하여서도 용어 정의, 요구사항, 참조 구조 등의 표준을 개발 진행 하고 있다. 현재 인공지능 표준화를 활발히 진행 중인 공적표준화 기구로는 ITU-T와 ISO/IEC JTC1의 SC42가 있으며, 사실표준화 기구로는 Khronos group과 IEEE가 대표적이다.¹⁾

국내에서는 한국정보통신기술협회(TTA)에서는 인공지능기반기술프로젝트 그룹(PG1005)를 운영하고 있으며, 지능정보기반 기술 중에서 인공지능의 기반기술과 관련된 국내 표준 개발을 담당하고 있다. PG1005는 인공지능의 기반기술표준 개발과 아울러 표준의 적합성, 상호운용성 및 시험 표준에 관한 연구 활동을 하고 있으며, 관련 국제기구와의 표준화 협력을 위해 필요한 활동을 포함한다.

〈표 4〉 TTA(한국정보통신기술협회)의 최근 제정된 국내 표준 예시

표준번호	표준명	제/개정일
TTAK,KO-10.1419	한국어 음성 및 텍스트 데이터의 의미적 정확성 품질검증 방법	2023-06-30
TTAK,KO-10.1420	컴퓨터 비전 분야 라벨데이터의 의미적 정확성 및 유효성 품질검증 방법	2023-06-30
TTAR-10.0174	OCR기반 차트 이미지 데이터 내용을 텍스트로 추출하는 인공지능 기술	2023-05-18
TTAK,KO-10.1414	지능형 질의응답 시스템을 위한 개체 연결 방법	2022-12-07
TTAK,KO-10.1341	지능형 질의응답 시스템 평가 체계	2021-12-08
TTAK,KO-10.1340	모의 환경 기반 인공지능 게임 에이전트 생성 요구사항	2021-12-08

4. 지속가능한 인공지능 개발

아직은 인공지능 관련 프로젝트가 계속적으로 운영되고, 지속적인 성능 향상과 유지보수가 이루어지는 경우가 드물다. 대개는 성능상 결함으로 인해 전체 프로젝트를 재설계하거나 학습 모델을 폐기하는 경우가 다수이고, 신뢰성 문제와 성능향상의 어려움으로 인해 서비스에도 장애요인이 있다.

MLOPS(Machine Learning Operations)는 데이터 관리 및 ML 시스템 개발과 서비스 운영(Operations)을 통합해 안정적으로 상품·서비스를 출시할 수 있도록 신속·유연한 개발을 추구하는 협업 방식으로 개발, 통합, 테스트, 출시, 배포, 인프라

1) 출처 : 인공지능 표준화 동향(한국전자통신연구원, 2019)

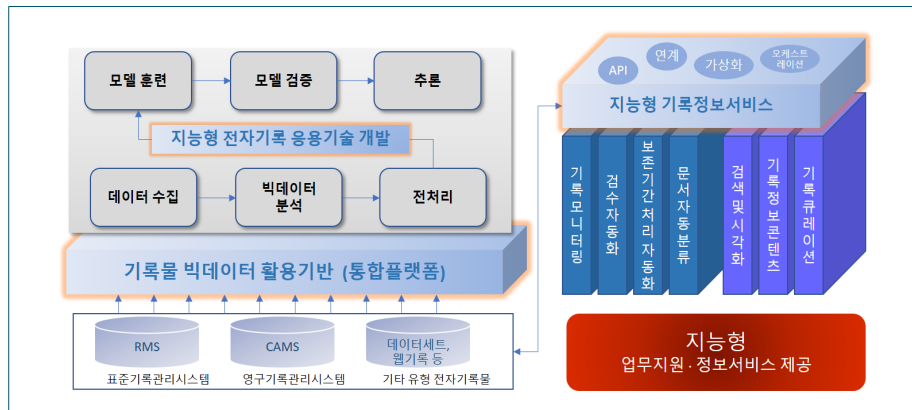
관리 등 ML 시스템 구성의 모든 단계에서 모델 재학습 자동화 및 모델 모니터링을 지원하며 ML 모델 라이프사이클을 관리한다. MLOps는 DEVOPS의 하위 개념으로, 만약 시스템 개발과 시스템 운영의 단절과 유지보수와 성능향상의 어려움이 상존한다면 도입을 고려해볼 필요가 있는 개념이다.

MS, AWS, 구글 등 글로벌 주요 IT기업들은 복잡한 절차가 필요한 AI·ML 기술개발의 효율성과 생산성을 향상하고자 MLOps 서비스를 개발하여 제공하고, 또 오픈소스로도 배포되고 있다. 국내에서도 많은 회사들에서 오픈소스를 기초로 한 제품들이 개발·출시되고 있으며 지속가능한 기계학습 개발을 위해 IT기업에서는 필수로 적용되고 있다.

IV. AI기술 관련 기록관리 연구개발

앞에서 기술한 내용들에 대한 조사·검토와 내부 의견수렴을 바탕으로 국가기록원에서는 지능형 기록관리 기술에 대한 연구개발계획을 수립하고 기술 적용방안 도출을 위해 노력하고 있다.

내부 이용자의 업무를 지원하는 기술과 외부 이용자의 정보서비스를 지원하는 기술로 나누어 국내외 사례 조사 및 분석을 수행하고, 최신 ICT 기술을 적용한 전자기록관리 자동화 기술개발 관련 기능분석을 통해 요구사항을 도출하는 등 인공지능기술의 기록관리분야 적용을 위한 탐색이 이미 여러차례 수행되었다. 비록 예산 상의 문제로 연구개발사업의 계획이 원안대로 추진되지는 못했지만 기존 국가기록관리 연구개발사업의 연구과제로 OCR 기술개발, 빅데이터를 이용한 자치단체 기록관리 추천 등 연구를 수행한 바 있다.



〈그림 8〉 지능형 기록관리기술 연구개발계획 및 기술개발 대상분야

2023년부터는 “국가기록관리·활용기술 연구개발 5개년 계획(2023-2027)”을 토대로 기록관리 통합플랫폼 개념 모델 설계 연구를 통해 디지털 전환 대응 지능형 전자기록관리 기반 구축을 시작한 상태이다. 이후 기록물 빅데이터 활용을 위한 기반기술의 확립과 기록 모니터링, 보존기간 관리, 문서분류, 기록물 언어모델 등 기록관리에 필요한 인공지능 기술을 적용할 계획이다.

또한 AI 학습용 데이터 구축 지원 사업을 통해 국가기록물 5,000철을 대상으로 3억 개의 말뭉치 데이터 및 5만 개의 질의응답용 지시데이터, 1만 개의 유해질의분류용 지시데이터를 구축하여 향후의 인공지능 도입을 위한 데이터 준비를 수행하고 있다.

향후 데이터 활용기반 강화와 지속적인 데이터 확보를 통해 단계적으로 기록물 관리에 특화된 지능형 요소기술 개발 및 시스템과의 상호 연계를 통한 지능형 업무지원·정보서비스 제공 방안 도출할 계획이다.

V. 맺음말

디지털 행정환경 변화 대응과 4차산업시대에서 중앙기록물연구관리기관으로서의 역할을 위해 인공지능 기술에 대한 적극적인 도입과 검토는 필수불가결한 일일 것이다. 다만 그 추진과정에서 기록관리 구성원간 AI 기술에 대한 공감, 기술 채택과 적용과정에서 지침·표준 활용의 중요성과 지속적인 개발을 위한 프로세스의 확립의 필요성을 본 글을 통해 강조하였다. 강조한 사항을 정리하면 다음과 같다.

- AI기술의 흐름과 데이터의 이해
- 문제 정의
- 기술의 완성도 검토
- 관련 표준 및 가이드의 활용
- 지속가능한 인공지능 개발

신기술과 기존 지식의 결합을 통해 대국민 서비스를 혁신하고 새로운 가치를 창출할 수 있고, 이는 최종적으로 기록관리 신속·효율화 및 대국민 기록서비스 확대·발굴할 수 있는 기회가 될 것이다. 기록관리분야 지능형 기술의 개발 및 적용에 대한 연구개발을 통해 광범위한 기록물 정보의 생산·검색 등 정보처리에 일관되고 예측가능한 접근법을 마련하여 효율적인 기록관리 및 기록정보의 활용이 확대되기를 기대하고 있다. 이를 위해

기술을 보유한 기관, 연구소, 기업 등과 협력하여 데이터 이용에 관한 생태계를 구축하는 것 또한 아카이브가 데이터 허브로서의 중추적 역할을 하는데 중요하다고 할 수 있다.

■ 참고문헌

- 빅데이터 · 인공지능 산업 진흥을 위한 데이터 과학의 발전 전략 연구(한국과학기술한림원, 2019)
- 공공분야의 인공지능 촉진 및 활성화를 위한 방안 연구(한국지능정보시스템학회, 2020)
- 책임있는 AI 채택 : 기업 AI 솔루션의 공공조달 가이드라인 (세계경제포럼, 2023)
- Perspectives and Prospects on Transformer Architecture for Cross-Modal Tasks with Language and Vision (Andrew Shin)
- 인공지능 표준화 동향(한국전자통신연구원, 2019)

|| 기록관리 이슈페이퍼 발간 목록 ||

발간호	제 목	작성자	발간일
vol. 1	기관 심층인터뷰를 통한 BRM 단위과제 운영 개선 방안 수립	황정원 기록연구사	2019. 10. 8.
vol. 2	『공공기록물법』 상의 기록의 개념 검토 ① 기록의 개념과 성립요건 - 정보와 증거로서의 기록의 함의를 기록물법에 적용하기 - ② 기록이란 무엇인가? - 『공공기록물법』에 따른 기록관리 대상의 범위와 관련하여 -	이젤마 서기관 임신영 기록연구사	2019. 10. 22.
vol. 3	대통령기록물 평가체계 개선 방안	윤정훈 행정사무관	2019. 10. 31.
vol. 4	“도전! 기록관리 명강사되기” 기록관리 강사양성제도 도입	김명옥 사서사무관	2019. 11. 15.
vol. 5	국가기록원 블록체인 기록관리 플랫폼 구축사업의 의미와 전망	왕호성 기록연구사	2019. 11. 22.
vol. 6	전자기록 장기보존정책의 방향	이지영 공업연구사	2019. 12. 5.
vol. 7	디지털기반 대통령기록관리체계 모델 재설계	김현숙 공업연구사	2019. 12. 12.
vol. 8	건축아카이브의 해의 동향 및 향후 과제 - ICAA BRAGA 2019 참가기 -	김수연 전문임기제 허인영 전문임기제	2019. 12. 13.
vol. 9	전자기록 장기보존패키지 모델 시험과 새로운 모델 제안	신동혁 공업연구사 김상국 전산사무관 나미선 학예연구관	2019. 12. 17.
vol. 10	기록물 매체수목 해의 동향 및 향후 과제	박지혜 공업연구관	2019. 12. 24.
vol. 11	기록물 생산현황 분석 결과(2018년 생산분)	김현애 기록연구사	2020. 1. 15.
vol. 12	기록물관리시스템을 통한 생산현황 통보 자동화 방안	하정하 기록연구관	2020. 1. 29.
vol. 13	기록물관리 전문요원 양성제도 현황과 전망 - 교육원 과정을 중심으로 -	성주영 기록연구사	2020. 2. 12.
vol. 14	정기실태점검을 통해서 본 기록관리 개선방안	박지태 학예연구관 송혜현 사서사무관	2020. 2. 27.
vol. 15	공공기록물법과 전자정보법과의 관계	임신영 기록연구사	2020. 3. 11.
vol. 16	해의 공공기록 평가선별제도 관련 사례 및 시사점	조영주 사서주사	2020. 3. 20.
vol. 17	속기록 의무생산회의에 대해 묻고 답하다	박이준 학예연구관 이주현 기록연구사	2020. 4. 3.
vol. 18	기록관리 현장 지원을 위한 기관방문 컨설팅의 추진	나창호 기록연구관 정경택 공업연구사	2020. 4. 21.
vol. 19	디지털화 기록의 문자인식 - OCR 적용 사례 및 테스트 결과를 중심으로 -	박지혜 공업연구관	2020. 5. 13.
vol. 20	이용 통계로 알아보는 국가기록포털의 현재	서경란 전산주사보	2020. 5. 27.
vol. 21	영구기록물관리기관의 비공개기록물 공개재분류 업무절차 개선	권미현 기록연구사	2020. 6. 17.
vol. 22	손상파일 유형과 복구사례로 본 전자기록의 이해	김자경 기록연구관	2020. 7. 21.
vol. 23	행정정보 데이터세트의 기록관리 방안 - 데이터세트 형태의 전자기록에 대한 기록관리 실행 방안 -	이주광 공업연구관	2020. 9. 11.
vol. 24	특수지의 특성과 기록보존	허인영 전문임기제	2020. 10. 23.
vol. 25	기록물 상태감사 개선방향과 데이터 활용 사례	박지혜 공업연구관 이상화 한시임기제	2020. 11. 20.
vol. 26	조선왕조실록 태백산사고본의 체계적인 보존관리와 개선방안	최현욱 학예연구사	2020. 12. 18.

|| 기록관리 이슈페이퍼 발간 목록 ||

발간호	제 목	작성자	발간일
vol. 27	중앙행정기관 기록관리기준표 개선 절차 수립 및 적용 사례 기록관리시스템 컨설팅 현황	황정원 기록연구사 이창영 공업연구관	2021. 2. 26.
vol. 28	대통령기록관 소장 미술품의 이해와 활용 영구기록물 분류체계 추진현황	김정은 학예연구관 김현진 기록연구사	2021. 3. 31.
vol. 29	공공기록물법령 질의답변 사례 - 전자기록물 관련을 중심으로 - 독립만세운동 기록물 복원·복제 지원 사례	김명옥 사서사무관 양소은 학예연구관, 조은혜 학예연구사	2021. 4. 30.
vol. 30	국가기록원의 기록관리 지원업무와 개선방향 민간분야 기록관리 관련 현황 및 시사점	문찬일 기록연구사 임신영 기록연구관, 이해원 기록연구사	2021. 7. 2.
vol. 31	디지털화 기록의 문자인식(OCR) 기술 적용 및 활용방안 기록관리 기관평가의 현재와 미래 디지털 사본의 원본인정을 위한 기준, 절차 및 관리방안 제안	홍정기 보건연구사 백승옥 사서사무관 김상국 전산사무관	2021. 9. 30.
vol. 32	지방기록물관리기관 설립동향 도입하고 싶은 해외기관의 기록물 보존정책	김지훈 사서주사, 최찬호 공업연구관 안규진 학예연구사, 이재영 인턴연구원	2021. 11. 30.
vol. 33	민간아카이브 기록정보 공유 플랫폼의 방향성을 위한 해외사례 분석 소셜미디어를 활용한 참여형 아카이브 전시기획	이정연 학예연구사 김주연 학예연구관	2022. 3. 31.
vol. 34	국제사회에서 아카이브 간 교류 협력 확대를 위한 노력 실효성 있는 기록관리 표준을 위한 단계적 정비 방안	임신영 기록연구관, 박민예 행정주사보 김명옥 서기관, 송현규 기록연구사	2022. 6. 30.
vol. 35	행정박물 선별·관리체계 개선방안 국가기록원 클라우드 전환 현황과 과제	박부숙 사서사무관, 박지영 사서주사보 왕호성 기록연구사	2022. 9. 30.
vol. 36	기록정보 열람서비스 과정에서의 기록관리 단계별 문제점과 개선방향 공동체 아카이브 관련 해외 동향 및 국내 현황과 과제 대통령 전자기록물의 이관방식 변경 및 평가	권미현 기록연구사 이정연 학예연구사 윤정훈 행정사무관	2022.12. 15.
vol. 37	국내외 기록관리 최신 동향 - 2022년 연구개발사업의 사례조사를 중심으로 -	김용일 공업연구사 이정연 학예연구사	2023. 4. 28.
vol. 38	2022년 국가기록원 국외업무출장 관련 해외 기록관리 사례	이젤마 서기관, 나미선 학예연구관 조은혜 학예연구사, 이정연 학예연구사 조이현 학예연구관, 신동혁 공업연구사 이상화 전문임기제 라급 이영락 공업연구사, 박성배 공업연구관	2023. 6. 30.
vol. 39	국내 기록관리 인공지능 기술 적용 검토	방재현 한국외국어대학교 정보·기록학과 겸임교수/(주)스토리안트 대표 김용일 공업연구사	2023. 9. 27

- 『기록관리 이슈페이퍼』는 기록관리 현장의 다양한 현안 논의와 기록인 여러분의 귀중한 연구성과 공유를 기다립니다. 기고를 원하시거나, 본지의 발간과 관련한 일체의 질의 혹은 건의사항이 있으실 경우, 혹은 이슈페이퍼 내용에 대한 질의 등 필자에게 전달한 사항이 있으신 경우는 뒷면의 문의처나 필자의 메일을 통해 전달하실 수 있습니다.

