

NLP 활용 사례 분석 및 도입에 관한 연구: 분석 프레임워크와 시사점*

박현정** · 임희석***

A Study on Use Case Analysis and Adoption of NLP: Analysis Framework and Implications

Hyunjung Park** · Heuseok Lim***

■ Abstract ■

With the recent application of deep learning to Natural Language Processing (NLP), the performance of NLP has improved significantly and NLP is emerging as a core competency of organizations. However, when encountering NLP use cases that are sporadically reported through various online and offline channels, it is often difficult to come up with a big picture of how to understand and interpret them or how to connect them to business.

This study presents a framework for systematically analyzing NLP use cases, considering the characteristics of NLP techniques applicable to almost all industries and business functions, environmental changes in the era of the Fourth Industrial Revolution, and the effectiveness of adopting NLP reflecting all business functional areas. Through solving research questions based on the framework, the usefulness of it is validated. First, by accumulating NLP use cases and pivoting them around the business function dimension, we derive how NLP techniques are used in each business functional area. Next, by synthesizing related surveys and reports to the accumulated use cases, we draw implications for each business function and major NLP techniques.

This work promotes the creation of innovative business scenarios and provides multilateral implications for the adoption of NLP by systematically viewing NLP techniques, industries, and business functional areas. The use case analysis framework proposed in this study presents a new perspective for research on new technology use cases. It also helps explore strategies that can dramatically improve organizational performance through a holistic approach that encompasses all business functional areas.

Keyword : Natural Language Processing(NLP), NLP Use Case, Artificial Intelligence(AI), Use Case Analysis Framework, NLP Adoption, Business Function

Submitted : February 4, 2022

1st Revision : February 19, 2022

Accepted : March 28, 2022

* 이 논문 또는 저서는 2020년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2020S1A5B5A16083616). 본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터지원사업의 연구결과로 수행되었음 (IITP-2018-0-01405).

** 고려대학교 Human-inspired AI 연구소 연구교수

*** 고려대학교 정보대학 컴퓨터학과 교수, 교신저자

1. 서 론

인류는 일상생활 언어인 자연어로 의사소통을 하며 다양한 문화와 문명을 축적해왔다. 오늘날과 같은 디지털 정보화 시대에는 스마트 디바이스, 사물인터넷(Internet of Things, IoT), 커넥티드카(connected car) 등으로 이루어진 초연결을 통해 인류의 의사소통 데이터가 폭발적으로 생성 및 저장되고 있다. 다양한 문헌 자료 및 지식 콘텐츠들이 디지털화되어 저장되기도 한다. 이러한 대규모 자연어 데이터 속에 숨어 있는 유용한 정보를 인간의 수작업으로 추출하여 오늘날의 리얼타임(real time) 비즈니스에 활용하는 것은 거의 불가능하다(Kang et al., 2020).

자연어처리(Natural Language Processing, NLP)는 컴퓨터가 자연어를 이해하거나 생성할 수 있도록 하는 학문 분야로서(임희석, 2019; Kang et al., 2020; Manning et al., 1999), 인간의 지각능력, 탐색능력, 추론능력, 학습능력, 자연어처리 능력 등을 소프트웨어로 구현하는 기술인 인공지능(Artificial Intelligence, AI)의 한 분야다. 최근 딥러닝(deep learning)이 NLP에 적용되면서 NLP의 성능이 크게 향상되었고 거의 모든 산업에 NLP가 도입되고 있다(임희석, 2019; Kang et al., 2020). 시장조사기관 모도 인텔리전스(Mordor Intelligence)는 세계 NLP 시장이 2020년 107억 달러에서 2026년 485억 달러까지 복합연간성장률(compound annual growth rate) 26.84%로 성장할 것으로 전망했다(Mordor Intelligence, 2020). 그리고 2021년 세계 기업들의 평균적인 NLP 예산은 전년도에 비해 10% 이상 증가한 것으로 조사되었다(Lorica and Nathan, 2021).

주지하는 바와 같이, 2019년 12월에 시작된 코로나19 팬데믹(COVID-19 pandemic)은 경제, 사회, 교육, 문화 전반에 걸쳐 비교적 서서히 진행되어오던 디지털 트랜스포메이션(transformation)에 박차를 가했다. 2020년 미국 소비자들은 오프라인 매장보다 전자상거래를 더 많이 이용했고, 2020년 11~12월 미국 소비자들의 온라인 쇼핑 지출액 또한

2019년 동기 대비 32% 급증한 역대 최대치를 경신했다(우미영, 2021). 짧지 않은 기간 동안 경험되고 학습된 언택트 라이프 스타일과 비즈니스 방식은 코로나 종식 후에도 뉴 노멀(new normal)로 지속될 가능성이 높다는 의견이 팽배하다(성경식, 2021; 우미영, 2021; Deloitte, 2021). 이에 따라, 비대면 고객 접점에서 차별화된 고객 경험을 제공하고 고객 만족도와 충성도를 높이는 고객 관계 관리 능력이 기업의 핵심 역량으로 부상했다(김영욱, 2021; 성경식, 2021; 우미영, 2021). 그리고 NLP는 이러한 변화에 대응하는 효과적인 솔루션으로서 주목받고 있다. NLP 기술은 고객 접점뿐만 아니라 기업 내부의 비즈니스 프로세스에도 혁명적인 변화를 일으키고 있다(Kang et al., 2020). 노르웨이 뉴스 에이전시인 NTB는 NLP 기술을 활용한 로봇틱(robotic) 저널리스트를 통해 스포츠 뉴스 기사 작성을 자동화함으로써 30초 만에 경기 결과를 보도할 수 있다고 밝힌 바 있다(Lever, 2019). 이제 기업들은 NLP를 활용하여 어떻게 내부 및 외부 고객의 시간, 금전, 노력 비용을 절약하고, 나아가, 휴머니즘이 가미된 인간적인 교류를 통해 부가 가치를 창출할 것인지를 고민하고 있다(Agarwal et al., 2021; Liu and Mazumder, 2021).

그런데 각종 온·오프라인 채널을 통해 산발적으로 보고되는 최신 AI 기술들의 활용 사례를 접하다 보면 기술의 종류가 너무 많고 변화 속도가 빨라 혼란스러울 때가 많다. NLP 기술로 범위를 좁혀도 종류가 정말 다양하다. 이러한 기술들을 어떻게 이해하고 해석할지에 대한 큰 그림(big picture)이 떠오르지 않거나 비즈니스로 연결하여 활용할 방법을 몰라 어려움을 느끼는 경우가 많다. 실제 비즈니스 현장에서 NLP 기술을 어떻게 활용해야 하는지에 대한 구체적인 시나리오 창출이나 도입 방안 수립에 어려움을 호소하는 기업들이 적지 않다.

대부분의 AI, 빅데이터 등의 신기술 활용 사례 관련 연구들은 몇 가지 대표적인 산업 또는 특정 산업의 활용 사례들을 자유로운 형식으로 소개한다(김동완, 2013; 김재생, 2014; 박아름 등, 2020; 복경수,

유재수, 2014; 이경주, 김은영, 2020; 임철수, 2015). 그런데 NLP 기술은 분석 대상이 자연어이고, 자연어는 거의 모든 산업 및 비즈니스 기능 영역에서 주요한 의사소통 수단으로 사용되고 있다. 이것은 NLP 활용 사례는 동종 산업뿐만 아니라 이종 산업 간 벤치마킹 가능성도 크다는 사실을 의미한다. 그리고 조직의 혁신적인 성과 향상을 위해서는 특정 비즈니스 기능 영역에만 NLP 기술을 도입하기보다는, NLP 적용이 가능한 모든 비즈니스 기능 영역을 아우르는 전체적인 관점에서의 접근이 필요하다는 것을 말해준다.

이와 같은 연구 배경을 토대로, 본 연구는 NLP 활용 사례들을 체계적으로 축적 및 분석하여 창의적인 비즈니스 시나리오를 생성하도록 지원함으로써 NLP의 비즈니스 활용도를 제고하려는 목적으로 출발했다. 그리고 이러한 연구 목적을 달성하기 위해 다음 세 가지 연구 문제를 도출했다.

- RQ1. NLP 활용 사례들을 체계적으로 축적 및 분석할 수 있는 적절한 프레임워크는 무엇인가?
- RQ2. NLP는 각 비즈니스 기능 영역에서 어떻게 활용될 수 있는가?
- RQ3. NLP 도입 현황 및 도입을 위한 시사점은 무엇인가?

2. 배경 지식 및 관련 연구

2.1 NLP 연구 및 분석 방법

NLP를 연구하는 패러다임은 대략 규칙 기반, 통계 기반, 기계학습 및 딥러닝 기반으로 변화해 왔다(임희석, 2019; Davenport, 2018; Kang et al., 2020). 첫째, 규칙 기반은 미리 정의해둔 언어의 문법적인 규칙에 기반하여 자연어를 처리하는 방식으로, 1950년대에 시작되었다. 언어 표현의 다양성으로 인해 정의해야 하는 규칙의 수가 너무 많고 예외가 항상 존재한다는 한계점이 있다. 둘째, 통계 기반

은 규칙 기반의 한계를 극복하기 위해, 임의의 언어 규칙은 문장 요소 사이의 유의미한 상관관계로 나타날 것이라는 아이디어로 1990년대에 시작되었다. 통계학의 ‘조건부 확률’ 개념을 NLP에 도입하여, 앞에 등장한 문장 요소에 대한 다음 문장 요소를 대규모 텍스트 데이터의 빈도 정보에 의해 확률적으로 선택하여 가장 자연스러운 문장을 생성한다. 셋째, 2000년대 이후의 기계학습 및 딥러닝 기반은 입력 데이터로부터 컴퓨터가 스스로 학습하여 문제 해결을 위한 가중치를 결정함으로써 모델을 만들게 하는 방식이다. 특히, 딥러닝 기반은 통계 기반 접근이 다루는 문장 요소간 등장 빈도 같은 일차적인 관계를 넘어 더욱 복잡하고 심층적인 관계까지 학습하는 장점이 있다.

NLP는 크게 자연어 이해(Natural Language Understanding, NLU)와 자연어 생성(Natural Language Generation, NLG)으로 구분할 수 있다. NLU는 입력 문장의 의미 및 의도를 정확히 파악하는 작업이고, NLG는 제공할 정보를 사용자 친화적인 자연스러운 문장으로 표현하여 출력하는 작업이다.

NLP의 고전적인 분석 단계는 (1) 불용어 제거, 띄어쓰기 및 맞춤법 교정 등의 텍스트 전처리(preprocessing), (2) 문자열을 최소 의미 단위인 형태소로 분리하는 형태소 분석(Morphological Analysis)과 문맥을 고려한 품사 태깅(Part-Of-Speech Tagging) 등의 어휘 분석(Lexical Analysis), (3) 문장의 문법 구조를 분석하여 구문 트리로 나타내는 구문 분석(Syntax Analysis), (4) 구문 분석에 의한 통사 구조에 따라 문장의 의미를 파악하는 의미 분석(Semantic Analysis), (5) 문맥 및 상황까지 고려하여 화자의 의도를 파악하는 화용 분석(Pragmatic Analysis) 순으로 진행된다.

최근 많은 NLP 작업(또는 기술) 영역에서 우수한 성능을 입증하고 있는 딥러닝 기반 분석 방법은 ‘사전훈련(pre-training)과 파인튜닝(fine-tuning)’이다(Devlin et al., 2018). 사전훈련은 레이블이 없는 대규모 데이터로 비지도학습(unsupervised learning)하여 언어모델이 일반적인 언어 지식을 습득하

도록 한다. 그리고 파인튜닝 단계에서는 사전학습된 언어모델을 감성분석 같은 특정 태스크(task)에 대한 레이블 데이터로 지도학습(supervised learning)하여 페러미터(parameter)를 조정함으로써 태스크 모델 성능을 올린다.

한편, 파인튜닝 없이 사전훈련만으로, 또는 하나나 몇 가지 예제를 통해 NLP 작업을 처리하는 방법도 소개되고 있다. 이러한 방법은 레이블 데이터가 부족한 도메인에도 적용할 수 있으며, 특히, NLG 작업에서 우수한 성능을 보인다는 사실이 보고된 바 있다(Brown et al., 2020).

2.2 NLP 적용 기회 발굴을 위한 벤치마킹

NLP와 같은 AI 및 IT(Information Technology)를 비즈니스에 적용할 시나리오를 창출하기 위한 방법으로, 가치 사슬 모델(value chain model)과 빅데이터 분석 과제 발굴의 하향식(top-down) 접근을 거론할 수 있다. 벤치마킹(benchmarking)은 이 두 가지에 모두 활용되며, 다른 조직들은 비즈니스 프로세스의 어떤 부분에 어떤 IT를 적용하여 어떤 성과를 달성하고 있는지에 대한 다양한 모범 사례들을 비교 검토하여 이루어진다.

가치 사슬 모델은 비즈니스 조직을 제품이나 서비스에 가치를 추가하는 본원적 활동들(primary activities)과 지원 활동들(support activities)의 시리츠로 간주한다(Laudon and Laudon, 2018). 본원적 활동은 제품이나 서비스의 생산 및 물류와 직접적으로 관계된 제조/생산, 판매/마케팅, 물류, 고객 서비스 등의 활동이다. 그리고 지원 활동은 본원적 활동을 가능하게 하는 인사/조직, R&D, 조달(procurement) 및 재무/회계, 계획, 품질 관리 등의 활동이다. 정보시스템은 비즈니스 조직이 원가 우위, 마케팅 믹스 차별화, 마켓 니치(niche) 집중화, 고객 및 공급자 친밀도 강화 전략 등을 통해 경쟁 우위를 달성하도록 지원한다. 가치 사슬 모델은 이러한 정보시스템의 전략적인 임팩트(impact)가 가장 잘 발휘될 수 있는 비즈니스 활동들이 무엇인지 구체적으로 규명하도록 도와준다

(Laudon and Laudon, 2018). 이때 다른 우수 조직이나 산업 베스트 프랙티스(best practice) 사례에 대한 벤치마킹이 활용된다(Laudon and Laudon, 2018). 가치 사슬의 범위는 상황에 따라 한 산업에 속한 협력사 및 고객 가치 사슬의 네트워크인 벨류 웹(value web)이나 서로 다른 산업에 속한 협력사들의 네트워크인 벨류 에코시스템(value ecosystem)으로 확장될 수 있다.

빅데이터 분석 과제 발굴의 하향식 방법은 비즈니스 모델을 거시적 환경 요인과 고객 및 투자자, 경쟁사, 협력사, 자사 역량 등을 고려하여 최적화하거나 비즈니스 컨텍스트(context) 변화에 따라 재설정하는 관점에서 빅데이터 기술의 적용 기회를 탐색하는 방법이다(한국데이터베이스진흥원, 2020). 이때 벤치마킹은 빠르고 쉽게 아이디어를 얻고 의미 있는 분석 기회를 포착할 수 있도록 도와준다(한국데이터베이스진흥원, 2020).

2.3 관련 연구

AI, 빅데이터, IoT 등 신기술의 활용 사례를 다룬 논문들이 있다(김동완, 2013; 김재생, 2014; 박아름 등, 2020; 복경수, 유재수, 2014; 이경주, 김은영, 2020; 임철수, 2015). 그런데 앞에서 언급한 바와 같이, 기존 연구는 대부분 몇 가지 대표적인 활용 사례들을 자유롭게 서술하는 형식으로 소개하고 있다. 본 연구처럼 다양한 활용 사례들을 체계적으로 축적 및 분석하는 프레임워크에 대해 논한 연구는 거의 없다.

활용 사례 관련 연구는 아니지만, 제조업의 산업 가치 사슬 단계에 따라 AI 적용 시나리오를 디자인할 때 사용할 수 있는 전반적인 레퍼런스 프레임워크를 제안한 논문은 있다(Zhang et al., 2019). 이들이 제안한 레퍼런스 프레임워크의 주요 차원에는 제조 산업의 가치 사슬 단계(application phase), 해당 단계의 AI에 대한 적용 요구사항(application requirement), 해당 단계의 적용 가능한 AI 지능 기술(intelligent technology), AI 지능 기술에 의해 실

현될 자기지각(self-perception)이나 자기통제(self-control) 같은 지능 기능(intelligent function) 등이 있다.

3. 연구 방법

3.1 연구 문제 해결 방법

서론에서 제시한 세 가지 연구 문제는 기본적으로 NLP 활용 사례를 분석하는 작업과 관련된다. 사례 분석 연구방법론은 대개 풍부한 관련 자료를 토대로 실제 현상의 맥락을 심층적으로 이해함으로써 어떤 경험적 현상이 왜 발생했고 어떤 과정을 통해 전개되었는지를 밝히는 연구에 적용된다(Curtis et al., 2000; Yin, 2003). 그런데 본 연구의 사례 분석은 조직의 NLP 도입 과정을 살펴봄으로써 시사점을 도출할 목적도 있지만, 다양한 활용 사례들을 통해 NLP 기술의 적용 기회를 발굴하고 비즈니스 시나리오에 대한 아이디어를 얻기 위한 목적이 더욱 크다. 사례연구의 설계, 자료 수집 준비, 자료 수집, 자료 분석, 사례연구 결과 보고 등 사례 분석 프로토콜의 요소 활동들은 이러한 목적을 반영하여 수행되어야 한다(Brereton et al., 2008; Yin, 2003).

이에 따라, 본 연구는 세 가지 연구 문제를 해결하기 위해 다음과 같이 접근한다. 먼저, 사례연구의 설계 및 자료 수집 준비 단계에서는, RQ1에 대한 대안으로서, 본 연구의 사례 분석 목적과 연구 문제를 반영하여 사례 추적 및 분석을 위한 프레임워크를 설계한다. 그리고 실제 사례를 추적하고 분석해보면서 필요한 차원을 추가하거나 불필요한 차원은 삭제하는 반복적인 과정을 통해 초기 프레임워크를 보다 적절하게 보완한다.

자료 수집은 데이터의 신뢰성을 확보하기 위해 출처를 논문, 신문 기사, 기관 보고서, 기관 웹사이트 및 블로그로 한정한다. 그리고 사례 서술의 구체성이 좀 부족하더라도 NLP 기술을 활용하는 아이디어 생성에 도움이 되는 사례라면 분석 대상으로 포함한다. RQ2는 도출된 프레임워크에 따라 추적된

NLP 활용 사례들을 비즈니스 기능 및 NLP 기술 차원을 중심으로 분석함으로써 해결된다. 이 과정을 통해, 도출된 프레임워크의 실용성도 입증된다.

RQ3을 해결하기 위해서는 활용 사례에 대한 정량·정성적 분석 외에 관련 설문조사, 보고서, 논문 등의 자료를 종합할 필요가 있다. 순수한 사례 기반 방법은 존재하지만 발표되지 않았거나, 발표되었지만 수집되지 않은 사례 등으로 인한 오차가 수반될 수 있다. 이러한 귀납적 접근의 한계점을 보완하기 위해 연역적으로 기술된 설문조사, 보고서, 논문 등을 종합하면, NLP 도입 현황을 좀 더 정확하게 파악하고 도입을 위한 시사점의 유용성을 높일 수 있을 것이다.

3.2 NLP 활용 사례 분석 프레임워크

본 연구에서는 모든 산업 및 비즈니스 기능 영역의 활용 사례들을 분석 대상으로 한다. 이것은 적용 범위가 넓은 NLP 기술의 특성, 산업간 경계가 희미해지는 4차 산업혁명 시대의 환경 변화(박문수, 이동희, 2017), 모든 비즈니스 기능 영역을 살펴보는 관점에서의 NLP 도입의 효과성 등을 고려하여 좀 더 혁신적인 비즈니스 시나리오 창출을 지원하기 위함이다. 하나의 NLP 활용 사례는 “어떤 산업의 어떤 비즈니스 기능 영역에서, 어떤 NLP 기술이 어떻게 활용되고 있는가”를 보여준다. 본 연구에서 이러한 NLP 활용 사례를 이해하고 추적 및 분석하는 프레임워크, 즉, RQ1에 대한 대안으로서 도출한 프레임워크는 [그림 1]과 같다. NLP 기술이 ‘어떻게’ 활용되고 있는가에 대응되는 차원이 ‘NLP 도입 이유’, ‘NLP 솔루션’, ‘NLP 도입 효과’이다. ‘국내/해외’ 차원은 국내와 해외 도입 현황을 비교하는 경우를 위해 추가하였다. ‘어떤 산업’, ‘어떤 비즈니스 기능 영역’, ‘어떤 NLP 기술’에 해당하는 차원에 대한 설명은 다음 절에서 좀 더 자세히 다룬다. 특정 산업이나 비즈니스 기능 영역, 또는 NLP 기술 영역별 분석은 분석 목적에 따라 [그림 1]의 프레임워크 차원을 피보팅(pivoting)하여 수행할 수 있다.

차원	산업 분류		비즈니스 기능 분류	NLP 도입 배경	NLP 솔루션	NLP 기술 분류	NLP 도입 효과	국내 / 해외
	대분류	중분류						
설명	(1) 1차 산업	농업/임업/수산업/광업	I. 제조/생산 II. 판매/마케팅 III. 재무/회계 IV. 인사/조직 * 제조/생산(판매/마케팅), 제조/생산(재무/회계), 제조/생산(인사/조직) 가능	NLP 기술을 도입하게 된 배경	당면 문제를 해결한 NLP 솔루션의 구현 과정 및 방법	주요 기술(정보추출/텍스트분류/자연어생성/기계번역/문서요약/...) 기타 기술(언어모델/개체명인식/토픽모델링/지식그래프/협업필터링/... * 솔루션에 사용된 NLP 기술	NLP 솔루션 및 기술 적용의 (비)재무적 효익	국내/해외 여부
	(2) 제조업	음식료품/자동차/전기장비/섬유가죽제품/화학제품/...						
	(3) SOC	전기가스/수도/하수폐기물/건설						
	(4) 서비스업	도소매/금융보험/보건복지/교육/공공행정국방/...						

[그림 1] NLP 활용 사례 분석 프레임워크

3.2.1 산업 분류

NLP 활용 사례와 해당 조직이 속하는 산업을 명시하기 위한 산업 분류 중 대분류는 (1) 1차 산업, (2) 제조업, (3) SOC(Social Overhead Capital), (4) 서비스업으로 나뉜다. 중분류는 이러한 대분류의 하위분류다. 예를 들어, 서비스업의 하위 중분류에는 도소매, 금융보험, 보건복지, 교육 등이 있다. 그리고 의료보험사 사례의 경우, 대분류는 ‘(4) 서비스업’이고 중분류는 ‘금융보험’ 도메인에 해당한다. 이러한 대·중 분류는 기본적으로 세계산업연관 데이터베이스(WIOD)를 바탕으로 하는 박문수, 이동희(2017)의 산업 분류를 채택하였다. 그리고 중분류 카테고리를 정하기 애매한 경우는 좀 더 자세한 내용이 제시된 제10차 개정 한국표준산업분류를 참고하였다.

3.2.2 비즈니스 기능 분류

NLP 활용 사례에 해당하는 비즈니스 기능은 I. 제조/생산, II. 판매/마케팅, III. 재무/회계, IV. 인사/조직으로 분류한다. 대부분의 비즈니스 조직은 이러한 비즈니스 기능에 의해 고객이 원하는 재화를 제조하거나 서비스를 생산하여 고객에게 제공한다. 네 가지 비즈니스 기능 중, I. 제조/생산 기능은 산업에 따라 고유한 프로세스 특성이 존재하지만, 나머지 세 가지 기능은 공통되는 부분이 많아 이중 산업간 벤치마킹 가능성이 더욱 커질 것으로 생각된다. 본 연구에서는 여러 산업의 사례들을 총괄하여

단순하게 분류하기 위해 제2.2절에서 설명한 일반적인 기업 가치 사슬의 물류, R&D, 조달, 품질 관리는 제조/생산 영역으로, 고객 서비스와 계획은 판매/마케팅 영역으로 매핑했다.

사례에 따라 자사 입장에서 제조/생산 기능이지만, 다른 조직이 벤치마킹하는 경우에는 판매/마케팅이나 인사/조직이거나, 제조/생산 외의 다른 기능이 혼합된 경우들이 있었다. 이런 경우에는 ‘제조/생산(판매/마케팅)’과 같은 형식으로 기록하였다.

3.2.3 NLP 기술 분류

본 연구에서는 다양한 활용 사례에서 사용될 수 있는 NLP 기술을 관련 서적 및 자료를 참고하여 <표 1>의 주요 NLP 기술 영역 7개와 Appendix에 있는 <표 A-1>의 기타 NLP 기술 영역 23개로 정리하였다. 주요 NLP 영역은 기타 NLP 영역에 비해 독립적으로 활용되는 경우가 더 많고, 기타 NLP 영역은 주요 NLP 영역의 컴포넌트로 사용되는 경우가 있다. 예를 들어, 개체명 인식(a-2)은 정보 추출(1)이나 텍스트 분류(2), 음성 인식(a-16) 및 음성 합성(a-18)은 질의응답(6)이나 대화시스템(7)의 구성 요소로 사용된다. 이러한 의존 관계는 주요 NLP 영역 간에도 존재한다. 예를 들어, 대화시스템(7)은 자연어 생성(3) 외에도 시스템 구현 방법에 따라 정보 추출(1), 텍스트 분류(2), 기계 번역(4), 문서 요약(5) 등의 기술을 모두 사용할 수 있다.

〈표 1〉 주요 NLP 기술 영역

NLP 기술	설명	분류
1. 정보 추출 (Information Extraction)	비정형 텍스트로부터 개체(entity)간 의미적 관계를 나타내는 튜플 형태의 구조화된 정보(ex. ‘고려대’, ‘in’, ‘서울’)를 자동 추출하는 기술로, 개체명 인식(NER), 공동참조규명(Co-reference Resolution), 관계 추출(Relation Extraction)로 이루어짐	NLU
2. 텍스트 분류 (Text Categorization)	감성분석, 의도(intent) 분류, 스팸 필터링, 뉴스 카테고리 분류 등과 같이 비정형 텍스트 문서를 해당 카테고리에 할당하는 기술	NLU
3. 자연어 생성 (Natural Language Generation)	문학 작품, 보고서, 이미지 캡션, 문서 요약, 번역, 대화 등의 자연어 산출물을 생성하는 기술	NLG
4. 기계 번역 (Machine Translation)	하나의 언어로 표현된 내용을 같은 의미를 나타내는 다른 언어로 자동 변환하는 기술	NLU/ NLG
5. 문서 요약 (Text Summarization)	장문의 텍스트에 대한 간결하고 정확한 요약을 생성하는 기술로, 추출요약(extractive summarization)과 추상요약(abstractive summarization) 방법이 있음	NLU/ NLG
6. 질의 응답 (Question & Answering)	사용자의 자연어 질문에 부합하는 답변을 문서로부터 찾아 제시하는 기술로, 정보 추출(IE), 정보 검색(IR) 등의 기술 필요	NLU/ NLG
7. 대화시스템 (Dialog System)	사람과 대화하는 컴퓨터 시스템으로, 챗봇(chatbot)이라 불림. 특정 목적 없이 넓은 범위의 주제에 대한 대화를 하며 심리적인 지원을 제공하는 챗봇(chit-chat) 시스템과 고객이 요청하는 정보를 제공하거나 과업을 수행하는 개인 비서 같은 목적 지향 시스템이 있음. 입출력 모드가 스피치인 경우 음성 인식, 자연어 이해, 대화 및 작업 관리, 자연어 생성, 음성 합성 등의 기술 필요	NLU/ NLG

4. NLP 활용 사례 분석

4.1 대표적인 활용 사례

비즈니스 기능 영역별로 대표적인 NLP 활용 사례를 소개한다. 이를 통해, 제3장에서 제시한 NLP 활용 사례 분석 프레임워크에 의해 활용 사례가 어떻게 이해될 수 있는지 확인할 수 있다. 동시에, 조직에서 NLP 기술이 어떻게 도입 및 활용되는지를 살펴볼 수 있다.

4.1.1 영국의 베네볼런트 AI(Benevolent AI)

- 산업 분류

대분류: (4) 서비스업 / 중분류: 연구개발(의료)

- 비즈니스 기능 분류

I. 제조/생산

- NLP 도입 배경

약물 재창출은 사용 승인을 받고 이미 사용되고 있는 약물이나, 출시 단계까지는 이르지 못했어도 임상시험에서 안전성이 확인된 약물 중 새로운 질병

에 효과적인 약물을 찾는 신약 개발 방법이다(박성은, 2021). 기존 신약 개발은 타겟 도출부터 개발 및 등록까지 10년 이상이 소요된다. 반면 약물 재창출에 의한 신약 개발은 일반적으로 3년 안에 가능하다. 그런데 긴박한 코로나19 팬데믹 상황에서는 3년이란 시간도 너무 길었다.

- NLP 솔루션

AI 기반 약물 발굴 기업인 영국의 베네볼런트 AI(www.benevolent.com)는 바리시티닙(Baricitinib)이 코로나19 치료제로 적합하다는 사실을 코로나19 팬데믹이 시작된 2020년 2월 처음으로 국제학술지에 발표했다(박성은, 2021). 연구팀은 바이러스가 세포 안으로 침투하도록 촉진하는 단백질인 AAK1의 활동을 억제하는 방법을 코로나19 치료법으로 제시했다. 기존의 378개 AAK1 활동 저해제 중에서 이미 사용 승인된 47개의 약물을 찾아내고, 이 중 AAK1과 결합력이 높은 6개의 저해제를 분자 구조 데이터와 의학 정보를 사용하여 선별했다. 그리고 이 중 부작용이 심하지 않고 투여 용량 대비 치료 효과가 큰 약물을 고른 결과가 바리시티닙이다.

코로나19 연구의 초기 단계에는 지식 그래프에 대한 패턴 쿼리 도구와 단백질-단백질 상호 작용 (PPI) 네트워크에 대한 분석 도구 같은 두 가지 인터랙티브 탐색 도구가 사용되었다(Smith et al., 2021). 지식 그래프에서 노드는 주로 생의학적인 질병, 단백질과 복합물, 질병 프로세스 등을 나타낸다. 그리고 노드 사이에 존재하는 인과 관계, 경로, 프로세스, 그룹 멤버십, 온톨로지 및 계층 등 다양하고 복잡한 생의학적인 관계는 링크로 표현된다. 이런 지식 그래프는 정제된 데이터베이스와 문헌 데이터에 대한 NLP 정보 추출 기술에 의해 구축된다. 그리고 새로운 생의학 문헌 자료에서 추출한 정보를 포함하도록 정기적으로 업데이트된다. 이러한 NLP 파이프라인을 통해 COVID-19와 같은 새로운 질병 관련 정보를 신속하게 추적하고 탐색할 수 있다.

- NLP 기술

1. 정보 추출, a-15. 지식 그래프

- NLP 도입 효과

2020년 11월 연구팀은 선정된 후보군에 대한 임상 시험을 진행, 유효성을 입증했다. 스웨덴 카롤린스카 의대와 영국 임페리얼 의대 공동연구팀은 바리시티닙을 코로나19 고위험군 중증환자 수십 명에게 투여한 뒤 생존율을 관찰했다. 결과적으로 바리시티닙을 투여한 환자가 투여하지 않은 환자에 비해 유의미하게 높은 생존율을 보였다. 2021년 7월 미국 식품의약국(FDA)은 본래 류마티스 관절염 치료제였던 바리시티닙을 코로나19 치료제로 정식 승인했다. 베네볼런트 AI의 NLP 기술은 신약 개발에 수반되는 비용과 실패 확률을 감소시키고, 신약의 타임투마켓(time to market)을 단축함으로써 의약품 산업의 판도를 바꾸고 있다.

4.1.2 미국의 에트나(Aetna)

- 산업 분류

대분류: 4. 서비스업 / 중분류: 금융보험

- 비즈니스 기능 분류

I. 제조/생산

- NLP 도입 배경

미국 의료보험사 에트나(www.aetna.com)가 심사하는 보험 청구 관련 자료는 재무 수치 데이터와 자연어 데이터가 자유로운 형식으로 혼합되어 있다. 그런데 이를 처리하는 소프트웨어는 의료기관마다 다른 언어 및 헬스 용어를 파싱(parsing)하는 성능이 좋지 않았다. 그래서 에트나는 직원 50명을 배치하여 각 계약별로 지급액, 공제액, 기타 수수료 등에 대한 설명을 읽고 보험금을 산정하여 청구 내역을 검토 및 정리하도록 했다. 에트나는 연간 240만 건의 계약을 처리해야 했기 때문에 이러한 작업은 몇 주에서 몇 달이 소요되었다. 게다가 보험금을 잘못 지급하는 경우도 종종 발생했고 이러한 오류는 고객들의 항의 전화로 이어졌다.

- NLP 솔루션

에트나는 2017년에 데이터 과학자, 설계자, 엔지니어, 개발자, 프로젝트 관리자 및 테스터로 구성된 프로젝트 팀을 구성했다. 그리고 전사 차원의 확장형 애자일 프레임워크(Scaled Agile Framework)와 반복적인 모델 개발 방법론을 적용하여 자동으로 보험 청구 관련 자료를 처리하는 앱을 개발했다. 프로젝트의 범위를 정하고 불량 및 중복 데이터를 처리하여 정제된 데이터 모델을 만든 후, 시스템 프로세스를 기술하고 성능을 평가하는 60개 이상의 코멘트 유형을 코딩했다. 그리고 78가지 언어의 애노테이터(annotator)를 개발, 훈련 및 테스트하고 금융 및 의료 비즈니스 도메인을 위한 130개의 사전을 정의했다. 완성된 보험 청구 처리 앱은 비정형 텍스트 파싱을 포함하는 NLP 기술과 스페셜 데이터베이스 소프트웨어를 결합하여 보험금 지급 관련 특성을 식별한다.

- NLP 기술

1. 정보 추출, a-2. 개체명 인식

- NLP 도입 효과

이렇게 개발된 보험 청구 자료 처리 앱으로 에트나는 IT Excellence 부문에서 2019 CIO 100 Award

를 수상했고, 연간 약 600만 달러의 심사 및 재작업 비용을 절감하고 있다. 의료기관들은 더 빠른 보상을 받게 되었고 처리 정확도 개선으로 불만 사항이 감소했다. 이러한 성공 사례는 애틀라가 새로운 비즈니스 유스케이스를 발굴하기 위해 AI 센터를 설립하는 계기가 되었으며 다른 AI 기술의 도입을 위한 발판 역할을 했다. 아직도 애틀라에는 단순 반복적인 작업들을 제거하고 자동화할 수 있는 많은 기회가 존재한다. 애틀라는 자동화를 사람의 일자리 대체보다는 사람의 일을 보조하는 관점에서 접근한다. 예를 들어, 보험 청구 심사를 담당했던 50명의 직원은 이제 더 높은 수준의 사고력과 더 많은 조율이 필요한 계약 및 보험 청구 건들에 집중하고 있다. 이 앱을 통해 애틀라는 의료 생태계에서 의료기관 및 환자들에게 더 나은 고객 경험을 제공하는 협력자가 되고 있다.

4.1.3 미국의 버라이즌(Verizon)

- 산업 분류

대분류: (4) 서비스업 / 중분류: 통신

- 비즈니스 기능 분류

II. 판매/마케팅

- NLP 도입 배경

미국의 통신 회사 버라이즌(www.verizon.com)은 전화나 이메일 또는 웹 포털을 통해 매달 10만 건이 넘게 접수되는 고객 요청을 처리해야 한다 (Stevie Awards, 2020). 고객 요청에 대한 응답 지연과 처리 오류는 고객 경험에 부정적인 영향을 미친다. 그래서 버라이즌은 가장 빈번한 유형의 고객 요청은 추론 및 탐색 방법을 사용하여 상황에 따른 대응 방식을 설정해주는 규칙 기반 시스템으로 자동화하여 처리하도록 했다. 그러나 그 효과는 제한적이었다. 고객 요청 내용의 문맥이나 의도 파악에 대한 정확도가 높지 않아 결국 엔지니어들의 참여가 빈번했고 투입 노동량이 별로 감소하지 않았다. 따라서 더욱 스마트한 시스템에 대한 요구가 절실해졌다.

- NLP 솔루션

버라이즌의 비즈니스 팀과 IT 팀은 함께 규칙 기반 시스템에서 발생하는 문제에 사용자 중심의 창의적 문제 해결 방법론인 디자인 씽킹(design thinking)을 적용했다. 그들은 딥러닝 기반 NLP 시스템이 해결책이 될 수 있다는 결론에 도달했다. 규칙 기반 시스템은 업데이트를 위해 코딩이 필요하고 정확도를 높이면 복잡성이 증가하므로 상당한 노력이 수반된다. 반면에 NLP 시스템은 적합한 훈련 데이터를 추가하기만 하면 보완이 가능하다. 버라이즌의 데이터 과학자들은 비즈니스 전문가들과 긴밀히 협력하여 모델을 구축하고, 모델을 훈련할 히스토리 데이터를 수집했다. 운영 엔지니어는 프로세스 초기에 모델 성능에 대한 피드백을 제공했다. 이와 같이, 신속하고 유연한 애자일 접근 방법(Agile approach)으로 2018년 1분기에 파일럿 모델을 선보였고 지속적으로 업그레이드하여 2018년 3분기에는 거의 모든 고객을 응대하는 글로벌 버전을 가동했다. 2018년 4분기에는 NLP를 활용하여 고객 요청의 대부분이 고객 요청 처리 진행 상태를 확인하는 것과 같은 단순 작업이고, 고객에게 제공한 응답 내용에 고객의 혼란을 야기하는 기술적인 전문 용어가 많이 포함되어 있다는 사실을 확인했다. 이에 따라, 딥러닝 기반 NLP를 활용하여 고객 요청을 더 쉬운 답변과 함께 더 빨리 해결할 수 있도록 시스템을 개선했다.

이렇게 탄생한 작품이 AI 디지털 워커(AI-enabled Digital Worker)였고, 버라이즌은 이로 인해 신기술을 혁신적으로 도입한 조직에게 수여하는 퓨처엣지50상(FutureEdge 50 Award)을 수상했다(CIO Online, 2021). AI 디지털 워커는 가장 흔한 고객 요청을 자동화하도록 문제 관리 시스템에 통합되어 운영되고 있다.

- NLP 기술

7. 대화시스템, 2. 텍스트 분류, a-2. 개체명 인식

- NLP 도입 효과

AI 디지털 워커는 2019년 2분기부터 매달 거의 1만 시간의 노동시간을 절약하고 있다. AI 디지털

워커가 대부분의 단순 반복적인 작업을 수행하므로 엔지니어들은 이메일 응답 같은 단순 작업에 시간을 소비하지 않고 좀 더 복잡한 네트워킹 문제를 처리한다. 버라이즌은 이제 고객 요청에 대해 99% 이상의 정확도로 몇 시간이 아닌 몇 초 만에 대응할 수 있다. 사람의 개입이 필요한 경우는 1% 미만이며, 모델을 지속적으로 재훈련하여 더 나은 성능을 달성해나가고 있다. 이러한 고객 서비스 응대 자동화를 위한 AI 디지털 워커 프로젝트는 버라이즌 혁신의 중추 모델로 활용될 예정이다.

4.1.4 독일의 도이치뱅크(Deutsche Bank)

- 산업 분류

대분류: 4. 서비스업 / 중분류: 금융보험

- 비즈니스 기능 분류

IV. 재무/회계

- NLP 도입 배경

대규모 펀드와 개인 투자자 모두 투자 의사 결정에 환경, 사회, 지배구조 관련 ESG 데이터를 반영하려는 수요가 증가하고 있다. 이러한 추가적인 고려 사항이 더 나은 포트폴리오 수익 창출에 도움이 된다는 증거도 있다. 대부분의 S&P 500 기업들은 지속가능성 보고서(sustainability report)를 발간하고 있는데, 비재무적인 ESG 데이터를 실행 가능한 정보로 변환하는 것은 까다로운 작업이다. 많은 기업들이 지속가능성과 관련된 방대한 자료를 제시하지만, 대부분 불투명하고 무의미하다.

- NLP 솔루션

도이치뱅크(www.db.com)는 대기업들이 보고서를 길게 작성할 수 있는 자원이 더 풍부하여 전반적으로 더 높은 ESG 등급을 받는 경향이 있다는 사실을 발견했다. 그리고 기업들이 진정으로 지속 가능한 방식으로 비즈니스를 수행하고 있는지를 측정하기 위해 지속가능성 보고서를 분석하는 대안적인 방법을 개발하기로 결정했다(FinText, 2020).

도이치뱅크는 탄소 배출을 줄이겠다는 기업들의

약속이 실제 지속가능성 성과와 관련이 있는지 조사했다. 보고서 내 탄소 관련 토론을 분석하면서 연구자들은 5가지 주제를 규명하고 각 주제와 관련된 상위 키워드를 선정했다. 기업들은 완화 및 적응 주제에 초점을 맞추어 순위가 매겨졌다. 이 시스템은 ‘첫 번째’와 ‘반’ 같은 숫자와 양적인 단어, 그리고 능동·수동적 언어를 분석했다.

- NLP 기술

a-9. 키워드분석, a-13. 토픽 모델링

- NLP 도입 효과

도이치뱅크는 매우 활동적이고 수치적인 언어를 사용하는 기업들이 미래의 배출량을 줄일 수 있는 가능성이 평균적으로 74%라는 사실을 발견했다. 또한, 기후변화 완화 또는 적응에 대해 자주 논의하는 기업은 실제로 감소 목표를 달성할 가능성이 65% 더 높았다.

4.1.5 스위스의 아테코(Adecco)

- 산업 분류

대분류: (4) 서비스업 / 중분류: 사업지원서비스

- 비즈니스 기능 분류

I. 제조/생산(아테코는 HR 솔루션 기업이므로 제조/생산 기능에 해당하지만 다른 조직에서는 IV. 인사/조직 기능에 적용 가능)

- NLP 도입 배경

스위스에 본사를 둔 아테코 그룹(www.adecgroup.com)은 포춘(Fortune)지 선정 글로벌 500대 기업으로, 인재파견, 아웃소싱, 헤드헌팅, 컨설팅 등을 제공하는 종합 HR 솔루션 기업이다. 전 세계 60개국에 5,100개 지사와 34,000명 이상의 내부 직원 네트워크를 보유하고 있으며, 매일 십만 개 이상의 고객사와 약 칠십만 명의 지원자들을 연결하고 있다. 아테코는 수많은 지원자들에 대한 정보를 처리하는 작업에 많은 시간과 비용이 소요되었다. 경쟁적인 포지션에 대한 일관성 없는 지원자 경험은 곧 높은 이탈률로 이어진다.

- NLP 솔루션

아테코는 리쿠르팅 업무의 효율성 및 효과성을 개선하기 위해 미국과 캐나다 전역의 창고, 콜센터 및 제조 포지션에 대한 지원자 서비스를 제공하는 마이아 리쿠르트(Myra Recruit)를 구현했다(Myra, 2021). 마이아(Myra)는 능동·수동적 지원자들을 모두 역동적으로 대화에 참여하도록 대화를 관리한다. 그리고 취업 관련 사이트를 방문한 잠재적 지원자들의 가장 시급한 질문에 답하거나 탐색 과정을 지원함으로써 더 많은 방문자들이 실제로 기업에 지원하도록 유도한다. 마이아는 비디오 기능도 제공한다.

- NLP 기술

7. 대화시스템, a-12. 정보 검색

- NLP 도입 효과

아테코는 매일 더 많은 지원자들에게, 더 많은 후속 조치를 제공하고 더 나은 홍보를 할 수 있었다. 마이아 개발자들은 단순히 지원을 요청하는 것보다 40% 더 많은 방문자들이 채용 프로세스에 실제로 참여하고, 마이아의 답변을 받은 지원자들은 채용 가능성이 200%까지 높아진다는 사실을 발견했다. 마이아 도입 후, 면접 시간 79% 단축, 채용자 생산성 123% 증가라는 성과도 달성하였다.

4.2 비즈니스 기능별 활용 사례 요약 및 시사점

본 연구에서 축적한 NLP 활용 사례는 156건이다. 산업 대분류에 의하면, 서비스업이 대부분이고, 제조업 3건, 1차 산업과 SOC가 각각 1건이다. 서비스업 안에서 내림차순으로 몇 가지 도메인을 열거하면, 금융보험 27건, 정보컴퓨터 18건, 공공행정국방 18건, 보건복지 15건, 도소매 14건, 교육 8건, 전문과학기술 6건, 연구개발 5건 등이다. 한편, 비즈니스 기능 분류에 의하면, 제조/생산 92건, 판매/마케팅 46건, 인사/조직 11건, 재무/회계 7건이다. 해외 사례가 많고 국내 사례는 20건이다.

본 절에서 다룬 비즈니스 기능별 활용 사례 요약과 시사점은 각각 RQ2와 RQ3에 관련되지만, 비즈

니스 기능별로 함께 제시하는 편이 더욱 효과적일 것으로 판단된다. RQ2에 대한 활용 사례 요약은 축적된 활용 사례들을 비즈니스 기능 및 기술 차원으로 피보팅하여 도출한 것이다. 산업에 따라 고유한 특성이 존재하는 제조/생산 비즈니스 기능 영역은 산업 대분류 차원을 추가적으로 사용하여 NLP 기술 및 산업 대분류에 따라 분석하였다. 그리고 나머지 비즈니스 기능 영역은 모든 산업에서 거의 그대로 적용할 수 있는 사례들이 대부분이므로 산업 대분류 차원을 사용하지 않았다. 또한, 질의 응답 기술은 질의 응답 챗봇처럼 사례를 통해서는 대화시스템과 구분하기 어려운 경우들이 많고, 실제로 목적 지향 대화시스템으로 구분하는 경우도 있어(Liu and Mazumder, 2021), 질의 응답과 대화시스템을 통합하여 정리했다.

4.2.1 제조/생산 비즈니스 기능

<표 2>는 제조/생산 기능의 NLP 활용 사례들을 NLP 기술 영역에 따라 정리한 내용이다. 네 가지 산업 대분류 중 NLP 기술이 가장 활발하게 적용되고 있는 분야는 서비스업이다. 서비스업에서는 보건복지(의료), 교육, 금융보험, 전문과학기술(법률), 연구개발, 도소매, 공공행정국방(공공행정), 정보컴퓨터 등의 도메인에서 대화시스템, 텍스트 분류, 정보 추출, 문서 요약, 자연어 생성 등의 NLP 기술을 적용하는 사례들이 증가하고 있다. 2021년 50개국 이상에서 655명의 기술 전문가들이 참여한 NLP 글로벌 온라인 서베이는 의료(healthcare), 컴퓨터전자기술(computers, electronics, and technology), 교육(education), 금융서비스(financial services) 순으로 대표 산업을 선정하였다(Lorica and Nathan, 2021). 모두 서비스업에 속하고 본 연구에서 축적한 사례 건수 분포와 비슷한 결과라는 것을 확인할 수 있다. 도메인별 기술 분포를 보면, 신약의 연구개발에는 정보 추출 및 지식 그래프와 문서 요약이, 금융보험에서는 감성분석 등의 텍스트 분류 사례가 많은 것처럼 도메인이나 업무 특성에 따라 다른 양상을 보인다.

〈표 2〉 제조/생산 비즈니스 기능 활용 사례 요약

NLP 기술		활용 사례 요약
정보 추출	1차 산업	- 기후, 토양 등의 지역적 조건을 반영하여 작물 재배 패턴을 결정함으로써 농업 생산성을 제고하는 NLP 기반 의사결정 지원 시스템[농업]
	서 비 스 업	- 보험료 청구 처리 자동화[금융보험] - 질병 및 단백질 등의 화합물 간 생의학적 관계 추출[연구개발(의료)] - 계약서 검토 및 분석[전문과학기술(법률)] - 다양한 언어/형식/스타일 문서에서 물품 정보 추출[항공운송]
텍스트 분류		- 정치인 사용 언어 및 표정 분석으로 감정 상태 파악[공공행정국방] - 에세이 과제 문법 및 문장 구조 반영 자동 평가/피드백, 수업 중 선생님/학생이 사용하는 언어를 분석하여 수업 몰입도 측정 및 교수법 개선[교육] - 기업 대출 신용평가모형에 반영될 수 있는 태도와 기업가정신 측정, 대출 프로세스 참여자들의 감정 분석, 분기별 경영실적 보고 내용의 기만 가능성 분석으로 투자 프로세스 보완, 대출 계약서 검토 및 사람 리뷰 필요 섹션 추출[금융보험] - 문제소지 학생의 트윗 감성분석으로 학교 총기 사고 예방[보건복지] - 법률 문서 분석 및 검토 필요 단락 추출[전문과학기술(법률)] - 뉴스피드 및 가짜 뉴스 필터링[정보컴퓨터]
	자연어 생성	- 에세이 과제 자동 평가/피드백[교육] - 제품설명 자동 생성[도소매] - 소셜 시나리오 작가[문화콘텐츠/기타] - 특정 토픽 관련 연구 논문 연대표 작성 및 여러 접근법/해결책 요약[연구개발] - 뉴스 작성[통신]
기계 번역		- 국방용 동시통역기[공공행정국방] - 여행용 동시통역기[문화콘텐츠/기타] - 다양한 언어 문서에서 물품 정보 추출[항공운송]
문서 요약		- 방대한 금융 연구 자료 요약으로 투자 전문가 및 포트폴리오 관리자 작업 지원[금융보험] - 특정 기준에 의한 신약 관련 연구 자료 요약[연구개발]
질의응답/ 대화시스템	제조업	- 퍼스널 스타일리스트 봇(고객이 올린 이미지를 컬러링하여 신발 디자인)[섬유가죽제품] - 가전제품, 자동차 챗봇 탑재[전기장비, 자동차]
	서 비 스 업	- 국민 생활법률 질문 맞춤형 답변, 세금관련 상담/민원처리/납부, 24시간 의학 DB 검색 맞춤 조언 국가의료서비스, 지역 의료서비스 연결, 음식안전 및 검사 관련 정보 제공, 정부기관 웹사이트 안내, 여권민원/자동차등록/주정차민원/시정일발 등 민발 민원 상담/처리[공공행정국방] - 수업내용 질문 답변 조교, 원투원 튜터링[교육] - 의료보험청구 처리 자동화, 불가능시 담당 직원 연결, 금융계좌 조회, 신용카드 관련 업무, 타 핀테크(FinTech) 서비스와 연계[금융보험] - 개인화 쇼핑/결제 서비스, 배송관련 정보 제공, 음성 주문 결제 기능[도소매] - 음악 재생[문화콘텐츠/기타] - 최신 논문, 과거 진료 정보 등의 정보를 스스로 학습하여 의사가 최적의 처방을 내리도록 보조, 개인 감정상태 관리 챗봇, 심리상담 및 인지행동 테라피[보건복지(의료)] - 택배 및 반품 예약[우편택배] - 투숙객 질문 답변[음식숙박] - 보증금회수, 출산휴가신청, 주차위반 등 법률서비스 챗봇[전문과학기술(법률)]
개체명 인식		- 개인 대출 관련 서류에서 날짜, 성별, 나이 등 신용평가모형에 들어갈 독립변수 값들을 추출[금융보험]
키워드 분석		- 기업 탄소배출 감량 이행도 예측 고려 지속가능 투자[금융보험]
문서 유사도		- 표절, 저작권 위반 탐지, 계약서 관련 조항 검색[전문과학기술(법률)]
정보 검색		- 관련 법령/판례 검색[전문과학기술(법률)]
토픽 모델링		- 기업 탄소배출 감량 이행도 예측 고려 지속가능 투자[금융보험]
지식 그래프		- 연관 법령 간 네트워크를 시각적으로 제시[전문과학기술(법률)] - 질병 및 각종 화합물 간 생의학적 관계 지식, 질병 정보 추출/축적/탐색[연구개발(의료)]
음성 인식		- 군장비 제어 음성 인터페이스[공공행정국방] - 동영상 자동 자막 생성[문화콘텐츠/기타]
음성 합성		- 책 저자 목소리 합성 및 낭독[문화콘텐츠/기타]

* 'NLP 기술' 열을 산업별로 분류하지 않은 행은 모두 서비스업에 해당함.

제조업 분야에서는 고객이 올린 이미지를 컬러링 하여 신발을 디자인하는 퍼스널 스타일리스트 챗봇 처럼 이미지 인식 기술과 NLP 기술을 함께 응용하는 사례가 보고되고 있다. 그리고 가전제품, 자동차 등에 챗봇을 탑재하는 등 제품 기능에 NLP 기술을 활용한 서비스를 융합하고 있다. 4차 산업혁명 시대에 급속히 성장하는 산업들이 제품과 서비스의 융합에 기반하고 있는 점을 고려할 때, 더욱 활발한 제조업과 서비스업의 연계 및 융·복합 노력이 기대된다(박문수, 이동희, 2017). 산업간 경계를 초월한 융·복합이 확대됨에 따라 과거에는 불가능했던 신산업 및 비즈니스 생태계가 창출될 것으로 예상된다.

한 산업의 도메인 지식이 다른 산업의 서비스 개선을 위해 사용되는 다음과 같은 사례도 존재한다. 미국에는 260 종류 이상의 비토종(nonnative) 진딧물이 있고 이로 인한 농작물 손실은 십억 달러에 달하는 것으로 추정된다(Parr et al., 2021). 미국으로 유입되는 모든 농작물에 대해 소리와 이미지 인식 기술을 적용할 수 있는데, 여기에 NLP 기술을 결합하면 해충 검사의 효율성과 정확도를 더욱 높일 수 있다. 방대한 디지털 문헌 자료에 대한 정보 추출로 특정 식물에 대한 해충 리스트를 도출하여 훈련 데이터에 포함되는 해충 이미지의 종류를 줄임으로써 모델 빌딩 시간을 단축할 수 있다. 또, 해충과 식물 특성을 모델 피쳐(feature)로 포함함으로써 더욱 성능이 좋은 모델을 구축할 수 있다(Parr et al., 2021).

그런데 <표 2>에서와 같이, 제조업은 서비스업에 비하면 아직 NLP 기술 수용도가 훨씬 낮은 것으로 보인다. 이것은 제조업의 비중 자체가 서비스업에 비해 현저히 낮거나 제조업의 특성상 NLP 기술의 유용성이 낮아서일 수 있다. 세계은행(The World Bank) 자료에 의하면, GDP 기준 2019년 미국의 서비스업 비중은 77.3%, 제조업은 10.9%, SOC는 10.9%, 1차 산업은 0.9%이고, 2020년 영국의 서비스업 비중은 72.8%, 제조업은 8.4%, SOC는 18.2%, 1차 산업은 0.6%이다(국가통계포털, 2021). 그리고 한국은행 자료에 의하면, 2020년 GDP 기준 한국의 서비스업 비중은 62.4%, 제조업은 27.1%, SOC

8.4%, 1차 산업 2%이다(국가통계포털, 2021).

세계적으로 볼 때, 한국의 산업 구조는 제조업 비중이 매우 높은 편이므로, 제조업의 NLP 수용도를 높이기 위한 노력은 파급 효과가 있을 것으로 예상된다. 예를 들어, 수많은 수송(shipment) 관련 문서를 개체명 인식 및 정보 추출을 통해 분석하여 공급 사슬(supply chain)의 지연 초래 단계를 파악하고, 자동화나 파이프라인 간소화 등으로 물류 효율성을 극대화할 수 있다. 그리고 해당 산업의 운송비용, 연료 가격, 노동 비용 등에 관한 온라인 정보 검색을 통해 산업 비율과 자사의 비율을 비교함으로써 비용 절감 기회를 포착할 수 있다. 또, 석유화학, 섬유 등의 연구개발에는 신약의 연구개발에 사용되는 지식 그래프 기술을 비슷하게 적용할 수 있을 것이다.

현재 대부분의 제조업은 서비스업과의 경계가 모호해지거나, 자동차 산업을 이끄는 3대 축인 전기차, 카셰어링(Car Sharing), 무인차가 전통적인 자동차 제조사가 아닌, AI 기업인 구글, 애플, 테슬라, 우버 등에 의해 주도되는 등 패러다임의 변화를 겪고 있다(Deloitte, 2021). 이러한 변화에 따라 제조업의 NLP 기술 수용도도 달라질 것으로 예상된다.

산업 대분류 중 비중이 제일 높은 서비스업에서 보건복지(의료), 교육, 금융보험 도메인의 제조/생산 비즈니스 기능에 대해 요약하면 다음과 같다.

• 보건복지(의료)

의료 도메인에서는 EHR(Electronic Health Record), 센서 데이터, 진단 자료, 관리 데이터 등 다양한 형태의 방대한 비정형 데이터가 생성된다. NLP 기술이 발전하면서 이러한 데이터 활용의 비즈니스 가치가 커지고 있다. 의료 도메인의 NLP 도입은 놀라운 성과를 달성했으며, 향후 5년간 가장 높은 연평균 성장률을 기록할 것으로 예측된 바 있다(Marketsandmarkets, 2021). NLP 기술이 몸과 마음의 질병을 진단하고, 효과적인 치료법을 추천하며, 중병 환자의 사망률을 예측하고, 치료 우선순위 및 임상 의사 결정을 지원하기 위해 활용되는 사례가 폭발적으로 증가하고 있다(Glaz et al., 2021). 앨라배마(Alabama) 대학교에서 수행한 한 연구에 의

하면, 의료 기록을 의사가 수작업으로 검토할 때보다 NLP 기법을 적용할 때 암 진단의 정확도가 22.6% 더 높게 나타났다(Osborne et al., 2016). 의료 관련 연구개발 도메인에서도 질병의 증상 추출, 심각성 분류, 유전자 역할 및 질병과의 상호작용 규명, 치료 효과 분석 등을 위해 NLP를 독립적으로 사용하

나 다양한 기계학습 방법론과 통합하여 사용하는 것이 의학 연구의 새로운 패러다임으로 고려되고 있다 (Glaz et al., 2021).

한 조사 연구에 의하면 이러한 의료서비스 분야 (medical field)에서 가장 널리 수행되는 NLP 작업은 정보 추출, 구문 분석, 텍스트 분류, 정보 검색,

〈표 3〉 판매/마케팅 비즈니스 기능 활용 사례 요약

NLP 기술	활용 사례 요약
텍스트 분류	- 고객 페르소나 파악, 잠재 고객(lead) 및 수익성 고객 정보 수집, 고객 대화 및 방문한 웹페이지 내용 분석을 통해 고객의 니즈 및 의도 분석, 서비스 개인화, 웹사이트 콘텐츠 개인화, 리얼타임 광고 전략 지원 - 고객 리뷰 및 피드백, 설문조사 리뷰 분석에 의한 고객 분류 및 고객 감정 원인 파악, 긍정/부정 이상의 슬픔/화/공포 등 고객의 구체적인 감정 파악, 브랜드 평판 등 VOC 데이터 감성분석, 광고 효과, 시장 상황 분석, 신기술/사건 등에 관한 뉴스 탐색 및 경쟁사 모니터링, 마케팅 우선순위 결정 - 플랫폼 관리(악성 댓글/코멘트 필터링)
자연어 생성	- 광고카피 자동 생성, 제품 페이지의 텍스트와 이미지 자동 추출하여 프로모션 동영상 제작
기계 번역	- 외국어 언어 사용 고객 시그먼트(segment) 공략, 장애인 고객 지원
문서 요약	- 중요 뉴스 및 자료 요약, 영업사원 대화 내용 요약
질의 응답 / 대화 시스템	- 고객 요청(질문/예약/주문/민원 등) 응답/처리/진행현황 안내 자동화 - 고객 사용 단어/말투/음성 고려 감성분석, 고객에게 유용한 정보 제공, 고객 관계관리, 고객 충성도 제고(24/7 이용 가능성) - 리드 생성(잠재 고객 정보 수집), 고객 맞춤 추천 및 큐레이션, 업로드 상품 이미지와 유사한 스타일 추천, 상황 인지 추천, 리얼 타임 세일 공지, 상품 혜택 등 마케팅 지원
개체명 인식	- 데이터에서 의사의 전문 분야, 지역 등 임상실험 관련 정보 추출 의사 선정, 의사가 환자들을 알므로 임상실험에 참여하도록 유도 - 제품명 및 기업명 인식, 제품 태소노미(taxonomy)에 의한 메타 데이터 자동 생성 및 제품 정보 자동 분류(웹사이트 탐색 정확도 향상 및 판매 데이터 분석 정확도 향상)
언어 모델	- 워드 벡터를 사용한 유사 제품이나 서비스 추천
키워드 분석	- 투자자 특성에 맞는 효과적인 마케팅 콘텐츠 작성 방법 분석(단어, 문장 길이 및 스타일, 토픽 등)
문서 유사도	- 시장 조사 설문지 작성시 질문 문항 라이브러리에서 시간이 지나도 변하지 않는 정보를 수집할 수 있는 최적 표현 문장 찾기(나이 vs. 출생년도) - 고객센터센터/웹프테스크에서 지식데이터베이스에 저장된 과거 케이스 중 유사도가 높은 케이스 선택으로 고객 대응시간 단축
토픽 모델링	- 고객 설문조사 주관식 문항 분석, 고객 불만 사항 분석, 고객 요구사항 분류 - 투자자 특성에 맞는 마케팅 콘텐츠 작성 방법 분석(단어, 문장 길이 및 스타일, 토픽 등)
협업 필터링	- 뉴스 등 콘텐츠 추천
지식 그래프	- 소셜 미디어 데이터에서 노드/링크에 해당하는 정보를 추출하여 구조화, 가장 많은 다갯 소비자에 도달할 수 있는 가장 적은 수의 인플루언서 선정
음성 인식	- 콜센터 대화 내용 기록

기계 번역 등이다(Wang et al., 2020). 그리고 <표 2>와 같이, 대화시스템도 최신 논문 및 학술 자료, 과거 진료 정보 등을 스스로 학습하여 의사가 최적의 처방을 내리도록 보조함으로써 진료 정확도 제고와 시간 및 비용 절감, 개인에 최적화된 맞춤형 케어를 실현하는 목적으로 개발 및 사용되고 있다. 또한, 심리상담을 하거나 질병 관련 정보를 수집 및 제공하는 목적으로도 활용되고 있다.

이외에도 의사의 임상 자료 입력 및 검색 효율성 개선, 환자와의 상호 작용 개선, 고품질 헬스케어에 대한 환자의 니즈 충족, E-Health를 통한 국민 건강 관리 지원 등을 위해서도 활용되고 있다(Locke et al., 2021).

• 교육

교육 도메인에서는 교사와 학생의 니즈에 따라 교육 서비스를 자동화 및 개인화하는 형태로 NLP가 활용되고 있다. 학생들의 텍스트 과제를 문법 및 문장구조를 반영하여 자동으로 평가하고 적절한 피드백을 제공한다. 그리고 웹 텍스트 데이터들을 활용하여 개인화된 교육자료를 생성하고 평가 문항을 자동으로 생성한다. 학생들의 질문에 대한 답변도 해준다. 궁극적으로는 비용 효과적이고 확장 가능한 대화시스템 기반의 원투원 튜터링 형태로 진화할 것으로 보인다(Litman, 2016). 정보 추출, 텍스트 분류, 자연어 생성, 질의 응답, 대화시스템 등의 NLP 기술이 많이 활용되고 있다.

• 금융보험

금융보험 도메인에서는 고객 보험 청구에 대한 처리 과정을 자동화하고, 사기 보험 청구나 돈 세탁 등의 이상 징후를 탐지하거나 정밀 조사 케이스로 분류하는 작업을 위해 NLP가 활용되고 있다(Chew, 2016; Fisher et al., 2016). 또한, 대규모 시장 조사 데이터를 분석하여 투자 위험을 진단하고, 경쟁사에 대한 분석에 기반하여 보험 상품을 설계하는 과정에도 이용된다. 이를 위해, 정보 추출, 텍스트 분류, 토픽 모델링, 대화시스템, 문서 요약, 개체명 인식 등의 NLP 기술이 활용되고 있다.

4.2.2 판매/마케팅 비즈니스 기능

<표 3>과 같이, 방대한 규모의 비정형 데이터 속에 감추어진 고객의 미충족 니즈(unmet needs)와 고통점(pain point)을 정확히 분석하고, 제품 및 서비스에 이를 반영함으로써 고객 만족도 및 충성도를 제고하기 위한 활동에 NLP 기술이 활용되고 있다. 고객의 제품 및 서비스 검색 효율성 및 효과성을 높이고, 리얼타임으로 파악한 고객 상황, 고객 프로필 및 구매 히스토리를 반영하여 개인화된 추천을 제공하는 사례들도 보고되고 있다. 최근 판매/마케팅 관련 연구에도 NLP 적용이 빠른 성장세를 보이고 있다(Kang et al., 2020). CRM 소프트웨어 회사들도 제품에 NLP 기술을 접목하고 있으며, NLP 기술을 활용하면 CRM을 향상시킬 수 있다는 사실이 보고된 바 있다(Ozan, 2021; Quarteroni, 2018).

판매/마케팅 기능의 NLP 기술 분포를 살펴보면, 텍스트 분류와 대화시스템의 활용도가 높으면서 토픽 모델링, 자연어 생성, 지식 그래프 등 다양한 기법들이 적용되고 있는 것으로 보인다. 특히, 일선 고객 지원(first line support)을 챗봇 대화시스템을 통해 자동화하여, 단순하고 반복적인 작업은 챗봇이 해결하고 복잡하거나 창의적인 능력을 요구하는 일은 사람이 처리하도록 하는 추세가 뚜렷하다. 챗봇은 24/7 이용 가능하므로 다양한 도메인에서 고객 만족도를 제고할 수 있는 효과적인 수단으로 떠오르고 있는 것이다. 다시 말하면, 챗봇은 도메인에 따라 구매자, 환자, 학생, 보험 가입자, 국민 등의 고객에 대한 접촉점에서 일대일 응대 서비스를 담당하도록 빠르게 확산되고 있다(한국정보화진흥원, 2018; Locke et al., 2021).

4.2.3 재무/회계 비즈니스 기능

<표 4>와 같이, NLP 기술은 재무제표 외에 대규모 SNS 데이터를 분석하여 기업의 재무적 성과 및 주가 예측의 정확도를 높이거나 디셉션(deception) 가능성을 분석함으로써 투자 의사 결정을 지원한다. 그리고 트랜잭션 이상 징후, 돈세탁 등 사기 가능성

을 탐지한다(Chew, 2016; Fisher et al., 2016). 또한, 비재무적 요소인 환경경영(Environmental responsibility), 사회책임경영(Social responsibility), 기업지배구조(Governance) 측면에서 윤리적인 책임을 다하는 기업에게 투자하도록 지원함으로써 ESG 패러다임 실현을 위해서도 활용된다. 즉, 기업들이 제출한 지속가능 경영보고서를 분석하여 목표 달성 가능성을 예측함으로써 지속가능 경영 성과가 높은 기업을 선별하는 것이다. 이러한 재무/회계 기능에서는 감성분석 같은 텍스트 분류, 토픽 모델링, 정보 추출 등의 NLP 기술이 많이 활용되는 것으로 보인다.

<표 4> 재무/회계 비즈니스 기능 활용 사례 요약

NLP 기술	활용 사례 요약
정보 추출	- 회계 감사시 잠재적 위험, 일별 트랜잭션 이상 징후, 돈세탁 등 사기 탐지
텍스트 분류	- 재무제표 외에 주가에 영향을 미치는 SNS 데이터 감성분석 - 긍정/부정 이상의 디셉션 관련 감성 분석 투자 결정 반영
자연어 생성	- 재무보고서 생성
키워드 분석	- 기업 탄소배출 감량 이행도 예측 고려 지속가능 투자
토픽 모델링	- 기업 탄소배출 감량 이행도 예측 고려 지속가능 투자

4.2.4 인사/조직 비즈니스 기능

<표 5>와 같이, 구직자 지원서 스크리닝 및 인터뷰 등 직원 채용 프로세스 효율화, 신입 사원 교육 자동화, 단순 반복적인 직원 업무 지원, 세금 및 회사 내규 관련 질의 응답 등의 활용 사례를 살펴볼 수 있다. 채용 담당자는 챗봇을 활용하여 구직자의 스피치, 표정 등을 분석하고 구직자의 업무 능력이나 적합도를 파악할 수 있고, 필요한 경우에 구직자에게 이메일을 보낼 수도 있다(Sengupta, 2021). 챗봇은 직원의 워크플로우를 디자인하고 회의 스케줄링, 회의실 예약, 세금이나 회사 내규에 대한 질의 응답 등 단순 반복적인 일을 대신할 수 있다. 직원들

은 부가 가치가 높은 일에 집중함으로써 업무의 질(quality)과 효율성 향상을 통해 직원 만족도를 증진시킬 수 있다. 그리고 챗봇은 회사 문화를 진단하는 직원 설문조사 문항을 만들고 개선해야 할 부분을 포착하도록 도와준다. 메신저, 모바일, 보이스 어시스턴트 등의 채널을 통한 커뮤니케이션 내용을 분석함으로써 내부 고객인 직원의 니즈를 파악하고 조직 몰입도도 측정할 수 있다. 신입 사원 훈련 후 안전 관리 조치 등에 대한 퀴즈를 실시하기도 한다.

<표 5> 인사/조직 비즈니스 기능 활용 사례 요약

NLP 기술	활용 사례 요약
텍스트 분류	- 구직자 지원서 스크리닝, 양질의 구직자 선별 - 스팸메시지, 봇 계정 등 필터링
기계 번역	- 장애인 직원 지원, 수화 통역 - 외국어 매뉴얼/카탈로그 번역, 실시간 회의 통역/번역 서비스, 카메라로 비추는 언어 즉시 번역
자연어 생성	- 이메일, 보고서, 설문조사 문항 생성
질의 응답 / 대화 시스템	- 구직자 면접시 직무적합도 판단, 커리어 사이트 유입 인원 실제 지원 전환 유도, 잠재적 지원자 질문 답변, 비디오 기능 가능 - 신입사원 교육/평가, 회사 내규 직원 질의 응답, 대화 내용 분석으로 직원 니즈 및 조직 몰입도 측정, 업무 스케줄 관리 - 서류 정리, 장부 작성 등 반복적이고 정형화된 업무 담당(직원들은 창의적이고 전략적인 업무에 집중)
문서 유사도	- 키워드 매칭보다는 의미 유사도 기반으로 희망 직무 적합도 산출
음성 인식	- 회의/통화 내용 기록 및 검색
음성 검색	- 회의 중에 원하는 회계 관련 자료를 언급만 해도 스크린에 보여줌

4.3 주요 NLP 기술 영역별 시사점

본 절에서는 RQ3과 관련한 추가적인 시사점을 제시한다. 본 연구에서 축적된 활용 사례들을 NLP 기술 영역 차원을 중심으로 피보팅하거나, <표 2>~<표 5>를 NLP 기술 영역을 중심으로 종합하여 살펴보면, 텍스트 분류, 대화시스템 및 질의 응답,

자연어 생성, 정보 추출 및 개체명 인식 등이 여러 비즈니스 기능에서 다양하게 적용되고 있는 것으로 보인다. 2021년 NLP 글로벌 온라인 서베이에서는 개체명 인식과 텍스트 분류가 가장 많은 활용 사례를 가지고 있고, 개체 연결(entity linking) 및 지식 그래프(knowledge graph)의 활용도는 AI 도입과 함께 증가하고 있는 것으로 나타났다(Lorica and Nathan, 2021). 그리고 질의응답과 자연어 생성 기술의 도입도 확대될 것으로 예측되었다. 이상의 내용을 종합하여 주요 NLP 기술 영역을 중심으로 기술 수용 현황의 특징을 정리해보면 다음과 같다.

첫째, 텍스트 분류는 전체 비즈니스 기능 영역에서 폭넓게 활용되는 NLP 기술이다. 특히, 감성분석을 살펴보면, 제조/생산 영역에서는 서비스업의 금융보험, 공공행정국방, 보건복지, 교육 도메인에서 활발하게 사용 및 연구되고 있는 것으로 보인다. 그리고 판매/마케팅 영역에서는 VOC 분석, 재무/회계 영역에서는 주가 예측 및 신용평가 등에 의한 투자 의사결정에 활용되고 있다. 이외에 문서의 다양한 카테고리 분류도 많은 활용 사례를 가지고 있다.

둘째, 개체명 인식을 포함하는 정보 추출과 문서 요약은 다양한 도메인에서 축적되는 방대한 텍스트 데이터로부터 유용한 지식을 효율적으로 추출해주는 NLP 기술로서, 빠르게 진화하는 지식정보사회의 핵심적인 기술이라고 할 수 있다. 보건복지(의료) 도메인에서 다양한 사례들이 보고되고 있고, 재무/회계 영역에서도 정보 추출과 문서 요약 활용 사례가 증가하고 있다. 그런데 판매/마케팅과 인사/조직 영역에서는 문서 요약 사례는 존재하지만 정보 추출 활용 사례는 부족한 것으로 보인다. 문서 요약은 NLP 도입 성숙(mature) 단계에 있는 기업에서 더 많이 활용되는 경향이 있다(Lorica and Nathan, 2021).

셋째, 자연어 생성은 도소매, 교육, 문화콘텐츠 도메인과 판매/마케팅, 재무/회계 영역에서 활용 사례들이 보고되고 있고 인사/조직 영역에도 도입이 확산될 것으로 예상된다. 광고 카피, 학생 과제 평가 의견, 뉴스 기사, 소셜 미디어 포스트, 소셜, 대시보드, 데이터 요약 보고서, 앱 통신, 재무 보고서, 이메

일 등 다양한 형태의 글을 작성하는 작업에 사용되고 있다. 콘텐츠 형식에 따라 구조화된 형태뿐만 아니라 시나 소셜 같은 창의적인 글쓰기도 가능하다. 자연어 생성을 도입 및 구현한 조직은 직원이 하나의 글을 작성하는 시간에 수천 개의 글을 생성할 수도 있다. 또한, 고객의 개인 정보를 반영하는 콘텐츠를 제작함으로써 섬세한 개인화를 통해 고객 경험을 개선하도록 사용될 수도 있다.

넷째, 질의 응답 챗봇을 포함하는 대화시스템은 여러 산업의 제조/생산 영역뿐만 아니라 판매/마케팅, 인사/조직 영역에서 도입 사례들이 증가하고 있다. 향후 몇 년 안에 모바일 챗봇은 마케팅 및 상거래 분야에 혁명을 일으킬 것으로 예측된 바 있다(Mordor Intelligence, 2020). 챗봇은 기업의 내외부 고객이 항상(24/7) 이용 가능하고, 방대한 지식 베이스를 효율적으로 검색하여 고객 질문에 대한 답을 찾아주고, 단순 반복적인 작업을 처리할 수 있다(한국정보화진흥원, 2018; Quarteroni, 2018). 따라서, 고객과 기업의 대화 채널로서 긍정적이고 효과적인 고객 경험을 저렴한 비용으로 제공할 수 있으며 기업의 업무 생산성 향상에 기여하고 있다. 고객의 감정과 대화 스타일을 반영하여 응대하고 대화 과정에서 필요한 정보를 추출하여 최상의 개인화를 위해 스스로 진화하도록 연구되고 있다(Agarwal et al., 2021; Liu and Mazumder, 2021; Pamungkas, 2019). 그런데 챗봇은 일부 기업사례에서는 비즈니스 가치를 창출하고 있지만, 많은 경우 서비스 범위와 유용성 및 편리성 면에서 고객이 기대하는 수준에 미달하여 아직 실질적인 이용률은 떨어지고 있는 것으로 보인다(한국정보화진흥원, 2018).

다섯째, NLP 애플리케이션의 입력 및 출력은 텍스트뿐만 아니라 음성, 이미지, 영상 등의 멀티모달 인터페이스를 지향하고 있다. 텍스트나 음성 인터페이스는 국제화 환경을 반영하여 다양한 언어를 지원하든가 애플리케이션 용도에 따라 지방 사투리를 사용하기도 한다. 특히, 음성 인터페이스는 기기 조작이 힘든 고령층의 경우나 차량 내부 환경에서의 편리함, 신원 인증 가능성, 음성이 담을 수 있는 감정

및 심리 정보 등의 장점으로 성장 가능성이 높다. 아직은 해결되어야 할 보안 이슈를 안고 있지만, 온라인 커머스와 모바일 커머스에 이어 음성으로 간편하게 물건을 주문하는 보이스 커머스의 잠재력 또한 스마트 스피커 및 스마트 TV의 보급과 함께 증가하고 있다(삼정KPMG 경제연구원, 2020).

4.4 기타 시사점

NLP 활용 사례들을 정성적으로 고찰함으로써 얻은 추가적인 시사점은 다음과 같다. 먼저, NLP 도입으로 인한 비즈니스 효과는 분명히 존재한다. 운영 비용을 절감하고 대기업이 가진 경쟁 우위를 신속하게 따라잡거나 새로운 서비스를 창출하도록 지원하기도 한다. 특히, 오픈 소스 소프트웨어는 NLP 도입 장벽을 낮추고 중소기업에게 새로운 가능성을 제공한다. 오픈 소스 챗봇 소프트웨어인 라사(Rasa)를 사용하여 유럽 시장에 5가지 언어로 서비스되는 고객 지원 챗봇을 구현하여 성공을 거둔 모바일 은행 N26의 사례가 존재한다(FinText, 2020). 문법 교정과 대화시스템 기술로 언어 교육 서비스를 제공할 수 있는 것처럼, NLP 기술을 응용하거나 조합함으로써 새로운 서비스가 창출되기도 한다.

다음으로, 성공적인 NLP 활용 사례는 기존 비즈니스 프로세스에서 처리해야 하는 텍스트 데이터나 단순 반복적인 작업의 양이 너무 많아 발생한 문제를 해결하기 위해 NLP 기술을 적용한 경우들이 많다. 업무 자동화를 통해 직원은 좀 더 고차원적인 일에 몰입함으로써 업무 생산성을 올릴 수 있다. 또한, NLP 솔루션은 대개 도메인 상황을 반영한 반복적인 조정 과정을 거쳐 구축되므로 성공적인 NLP 도입을 위해서는 NLP 기술과 함께 도메인 지식이 필요하다. 일반적인 언어모델이 아닌 도메인에 특화된 언어모델을 사용하기도 한다. 중국의 최대 전자상거래 기업인 알리바바(Alibaba)와 같이, 자체적으로 개발한 언어모델을 전자상거래 플랫폼의 고객 서비스 챗봇, 검색 엔진, 헬스케어 데이터 분석 등 알리바바의 비즈니스 생태계에 광범위하게 적용하는 사

례도 존재한다(Alibaba Clouder, 2020).

5. 결 론

본 연구에서는 거의 모든 산업 및 비즈니스 기능에 적용할 수 있는 NLP 기술의 특성을 고려하여 NLP 활용 사례들을 체계적으로 이해하고 분석할 수 있는 프레임워크를 제안했다. 그리고 이를 토대로 연구 문제들을 해결함으로써 그 실용성을 입증했다. 먼저, NLP 활용 사례들을 축적하여 비즈니스 기능 영역 및 NLP 기술 차원을 중심으로 피보팅하여 분석함으로써 NLP 기술이 각 비즈니스 기능 영역에서 어떻게 활용되고 있는지를 도출했다. 제조/생산 비즈니스 기능 영역에서는 산업 대·중분류 차원에 의해 산업 및 도메인별 분석도 수행할 수 있음을 보였다. 그리고 제조/생산 영역의 특정 산업 및 도메인 활용 사례들을 판매/마케팅, 재무/회계, 인사/조직 영역의 활용 사례들과 연결하면 어떤 NLP 기술이 해당 산업 및 도메인의 비즈니스 기능 영역 전반적으로 어떻게 활용될 수 있는지를 살펴볼 수 있음도 제시했다. 또한, 활용 사례 외에 관련 설문조사, 보고서, 논문 등의 자료를 종합하여 비즈니스 기능 및 주요 NLP 기술 영역별 시사점을 도출했다.

본 연구는 다음과 같은 한계점을 고려하여 이해될 필요가 있다. 첫째, NLP 활용 사례는 존재하지만 수집되지 않아 분석에서 제외된 경우들이 있다. 검색 행위에 따라 수집 여부가 달라질 수 있고 시간이나 노력 비용의 한계도 있었다. 둘째, NLP 활용 사례 분석 프레임워크에서 사용한 NLP 기술 분류는 MECE(Mutually Exclusive and Collectively Exhaustive) 조건을 충족하지 않는다. 이것은 각각의 컴포넌트 기술들도 독립적으로 활용되는 경우가 다수 존재하기 때문에 이에 대한 사례를 수집하기 위해서는 허용해야 하는 부분이었다. 셋째, NLP 기술을 명시한 사례는 그대로 기록했지만, 구체적으로 언급하지 않은 사례는 NLP 도입 배경, 솔루션, 도입 효과에 서술된 내용에 기반하여 개연성이 높은 좀 더 포괄적인 NLP 기술을 추론하여 기록하였다. 따

라서, 의존적이거나 부수적인 NLP 기술은 누락 되었을 가능성이 존재한다.

이러한 한계점에도 본 연구는 NLP 기술과 산업 및 비즈니스 기능 영역을 체계적으로 조망함으로써 혁신적인 비즈니스 시나리오 생성을 촉진하고 NLP 도입을 위한 유용한 시사점을 제공한다. 본 연구에서 제안한 활용 사례 분석 프레임워크는 신기술 활용 사례 관련 연구를 위한 새로운 관점을 제시한다. 또한, 모든 비즈니스 기능 영역을 아우르는 전체적 접근(holistic approach)을 통해 조직의 성과를 전반적이고 획기적으로 향상시킬 수 있는 전략을 모색하도록 지원한다. 어떤 NLP 기술이 전체 비즈니스 기능 영역에서 어떻게 활용될 수 있는지에 대한 정보는 NLP 도입 및 시스템 구축 전략 수립에서 중요한 역할을 할 수 있다.

본 연구는 NLP 기술을 중요하게 다룬 측면이 있다. 그런데 파급적 혁신이 성공하기 위해서는 디지털 기술도 중요하지만 결국 고객이 느끼는 가치가 핵심일 것이다. 고객을 중심으로 제품과 서비스를 다시 바라보고 비즈니스 프로세스와 모델을 혁신하는 노력이 필요하다. 다양한 활용 사례에 대한 체계적인 벤치마킹을 지원하는 본 연구가 이러한 노력의 결실을 풍성하게 하는 자양분이 되길 바란다. 그리고 아직 모든 비즈니스 기능 영역에 NLP를 도입한 조직의 사례는 보고되지 않았지만, 이러한 성공 사례를 접해볼 수 있기를 기대한다.

참고문헌

- 과학기술정보통신부, 정보통신산업진흥원, *클라우드 산업실태조사 결과보고서*, 2020.
- 국가통계포털, “경제활동별 국내총생산(OECD)”, 2021, https://kosis.kr/statHtml/statHtml.do?orgId=101&tblId=DT_2KAA906_OECD (Accessed on Nov. 15, 2021).
- 김동완, “빅데이터의 분야별 활용사례”, *경영논총*, 제34권, 2013, 39-52.
- 김영욱, “고객의 비즈니스를 이해하는 대화형 AI”, *한국인터넷진흥원(KISA)*, 2021.
- 김재생, “빅데이터 분석 기술과 활용사례”, *한국콘텐츠학회지*, 제12권, 제1호, 2014, 14-20.
- 박문수, 이동희, “4차 산업혁명 시대 주요국 제조업과 서비스업 연계성 현황과 시사점”, *산업연구원 ISSUE PAPER*, 2017.
- 박성은, “코로나19로 시급해진 신약 개발, AI가 유망 해결사”, *AI타임즈*, 2021년 10월 1일.
- 박아름, 이새봄, 송재민, “인공지능 기반 챗봇 기술의 산업 적용 연구”, *한국컴퓨터정보학회 논문지*, 제25권, 제7호, 2020, 17-25.
- 복경수, 유재수, “빅데이터 활성화 정책 및 응용 사례”, *정보과학회지*, 제32권, 제11호, 2014, 46-57.
- 삼정KPMG 경제연구원, “음성 AI 시장의 동향과 비즈니스 기회”, *Issue Monitor*, 제126호, 2020년 4월.
- 성경식, “디지털화한 ‘뉴노멀’ 시대에 인간성 교감 더하기”, *전자신문*, 2021년 1월 14일.
- 우미영, “코로나가 바꾼 고객경험의 뉴 노멀”, *Adobe Blog*, 2021년 3월 2일.
- 이경주, 김은영, “플랫폼 서비스 혁신에 있어 인공지능(AI)의 역할과 효과에 관한 연구: 카카오 그룹의 인공지능 활용 사례 연구”, *지식경영연구*, 제21권, 제1호, 2020, 175-195.
- 임철수, “IoT 서비스 활용사례 분석 및 산업 활성화 이슈”, *한국차세대컴퓨팅학회 논문지*, 제11권, 제6호, 2015, 41-50.
- 임희석, 고려대학교 자연어처리연구실, *자연어 처리 바이블*, 휴먼 사이언스, 2019.
- 한국데이터베이스진흥원, *데이터 분석 전문가 가이드*, 한국데이터산업진흥원, 2020.
- 한국정보화진흥원, “인공지능 기반 챗봇 서비스의 국내외 동향분석 및 발전 전망”, *Trend & Future*, 2018-2호, 2018.
- Agarwal, A., S. Maiya, and S. Aggarwal, “Evaluating Empathetic Chatbots in Customer Service Settings”, 2021, arXiv:2101.01334 [cs.CL].
- Alibaba Clouder, “Natural Language Intelligence:

- Building a Language Bridge for Business”, *Alibaba Cloud Blog*, 22 July 2020.
- Brereton, P., B.A. Kitchenham, D. Budgen, and Z. Li, “Using a protocol template for case study planning”, *12th International Conference on Evaluation and Assessment in Software Engineering (EASE)*, Vol.8, 2008, 41–48.
- Brown, T.B., B. Mann, N. Ryder et al., “Language Models are Few-Shot Learners”, 2020, arXiv:2005.14165 [cs.CL].
- Chew, P. A., “Natural Language Processing Meets Accounting, Increases Accuracy, and Reduces Costs”, Report GCG003, *Galisteo Consulting Group*, 2016.
- CIO Online, “FutureEdge 50 Awards 2021”, 2021, <https://app.qwoted.com/opportunities/award-futureedge-50-awards-2021> (Accessed on Nov. 7, 2021).
- Curtis, S., W. Gesler, G. Smith, and S. Washburn, “Approaches to sampling and case selection in qualitative research: Examples in the geography of health”, *Social Science & medicine*, Vol.50, No.7–8, 2000, 1001–1014.
- Davenport, T.H., “From analytics to artificial intelligence”, *Journal of Business Analytics*, Vol.1, No.2, 2018, 73–80.
- Deloitte, *Deloitte Insights Magazine*, No.17, 2021.
- Devlin, J., M. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”, 2018, arXiv:1810.04805 [cs.CL].
- FinText, *8 Case Studies in Banking and Investment Management*, 2020.
- Fisher, I.E., M.R. Garnsey, and M.E. Hughes, “Natural Language Processing in Accounting, Auditing and Finance: A Synthesis of the Literature with a Roadmap for Future Research”, *International Journal of Intelligent Systems in Accounting and Finance Management*, Vol.23, No.3, 2016, 157–214.
- Friedman, C., “A broad-coverage natural language processing system”, *Proceedings of AMIA Symposium*, 2000, 270–274.
- Glaz, A.L., Y. Haralambous, D. Kim-Dufor, P. Lenca, R. Billot, T.C. Ryan, J. Marsh, J. DeVyllder, M. Walter, S. Berrouguet, and C. Lemey, “Machine Learning and Natural Language Processing in Mental Health: Systematic Review”, *Journal of Medical Internet Research*, Vol.23, No.5, 2021, e15708.
- Kang, Y., Z. Cai, C. Tan, Q. Huang, and H. Liu, “Natural language processing (NLP) in management research: A literature review”, *Journal of Management Analytics*, Vol.7, No.2, 2020, 139–172.
- Laudon, K.C. and J.P. Laudon, *Management Information Systems: Managing the Digital Firm*, Pearson, 2018.
- Lever, R., “Robo-journalism gains traction in shifting media landscape”, *AFP News*, 10 March 2019.
- Litman, D., “Natural Language Processing for Enhancing Teaching and Learning”, *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol.30, No.1, 2016.
- Liu, B. and S. Mazumder, “Lifelong and Continual Learning Dialogue Systems: Learning during Conversation”, *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol.35, No.17, 2021, 15058–15063.
- Locke, S., A. Bashall, S. Al-Adely, J. Moore, A. Wilson, and G. B. Kitchen, “Natural language processing in medicine: A review”, *Trends in Anaesthesia and Critical Care*, Vol.38, 2021, 4–9.
- Lorica, B. and P. Nathan, “2021 NLP Survey”,

- Gradient Flow*, 2021.
- Manning, C.D., C.D. Manning, and H. Schütze, “Foundations of Statistical Natural Language Processing”, *MIT Press*, 1999.
- Marketsandmarkets, *Natural Language Processing Market - Global Forecast to 2026*, 2021.
- Mordor Intelligence, *Natural Language Processing (NLP) Market - Growth, Trends, COVID-19 Impact and Forecasts (2021-2026)*, 2020.
- Myra, “Customer Stories”, 2021, <https://www.myra.com/customer-stories/> (Accessed on Nov. 28, 2021).
- Osborne, J., M. Wyatt, A. Westfall, J. Willig, S. Bethard, and G. Gordon, “Efficient identification of nationally mandated reportable cancer cases using natural language processing and machine learning”, *Journal of the American Medical Informatics Association*, 2016, 1077-1084.
- Ozan, Ş., “Case studies on using natural language processing techniques in customer relationship management software”, *Journal of Intelligent Information Systems*, Vol.56, 2021, 233-253.
- Pamungkas, E.W., “Emotionally-Aware Chatbots: A Survey”, 2019, arXiv:1906.09774 [cs.CL].
- Parr, C.S., D.G. Lemay, C.L. Owen, M.J. Woodward-Greene, and J. Sun, “Multimodal AI to Improve Agriculture”, *IT Professional*, Vol.23, No.3, 2021, 53-57.
- Quarteroni, S., “Natural Language Processing for Industry”, *Informatik Spektrum*, Vol.41, 2018, 105-112.
- Sengupta, R., “How Natural Language Processing can Revolutionize Human Resources”, 2021, <https://www.aihr.com/blog/natural-language-processing-revolutionize-human-resources/> (Accessed on Nov. 17, 2021).
- Smith, D.P., O. Oechsle, M.J. Rawling, E. Savory, A.M.B. Lacoste, and P.J. Richardson, “Expert-Augmented Computational Drug Repurposing Identified Baricitinib as a Treatment for COVID-19”, *Frontiers in Pharmacology*, Vol.12, 2021, 709856.
- Stevie Awards, “Verizon - Best Technical Support Solution”, 2020, <https://stevieawards.com/aba/verizon-best-technical-support-solution> (Accessed on Dec. 10, 2021).
- Wang, J., H. Deng, B. Liu, A. Hu, J. Liang, L. Fan, X. Zheng, T. Wang, and J. Lei, “Systematic Evaluation of Research Progress on Natural Language Processing in Medicine Over the Past 20 Years: Bibliometric Study on PubMed”, *Journal of Medical Internet Research*, Vol.22, No.1, 2020, e16816.
- Yin, R.K., *Case Study Research: Design and Method*, Newbury Park, CA: SAGE Publication, 2003.
- Zhang, X., X. Ming, Z. Liu, D. Yin, Z. Chen, and Y. Chang, “A Reference Framework and Overall Planning of Industrial Artificial Intelligence (I-AI) for New Application Scenarios”, *The International Journal of Advanced Manufacturing Technology*, Vol.101, 2019, 2367-2389.

〈Appendix〉

〈표 A-1〉 기타 NLP 기술 영역

NLP 기술	설명
a-1. 언어모델 (Language Model)	언어를 이루는 구성 요소인 글자, 형태소, 단어, 문장, 문단 등의 구성 요소에 확률값을 부여하여 이를 바탕으로 다음 구성 요소를 예측하거나 생성하는 모델로서, 자연어 생성, 기계 번역, 문서 요약, 질의 응답, 대화시스템, 음성 인식 등 다양한 NLP 태스크 모델의 기반이 됨
a-2. 개체명 인식 (Named Entity Recognition)	비정형 텍스트에서 사람, 장소, 기관, 날짜, 또는 도메인에 따라 약물, 단백질 등 고유한 의미의 명명된(Named) 개체(Entity)를 식별(Recognition)하는 기술로, 질의 응답, 정보 검색(Information Retrieval), 관계 추출(Relation Extraction) 등의 구성 요소임
a-3. 띄어쓰기 교정 (Spacing Correction)	텍스트 전처리 단계에서 띄어쓰기를 교정하는 기술로, 규칙/통계/딥러닝 기반 방법이 있음
a-4. 맞춤법 교정 (Spelling Correction)	텍스트 전처리 단계에서 텍스트에 포함된 맞춤법 오류를 감지하고 수정하는 기술로, 정확한 의미 전달 및 정보 교환을 위해 필요하며, 규칙/통계/딥러닝 기반 방법이 있음
a-5. 형태소 분석 (Morphological Analysis) / 품사 태깅(Part-Of-Speech Tagging)	어휘분석(Lexical Analysis) 단계에서 최소 의미 단위의 형태소를 사용하여 단어 형성 방법을 규명하는 기술로, 단어를 형태소로 분리하기, 형태소 연결 시 일어나는 변형 전의 원형 찾기, 문맥을 고려한 형태소-품사 페어 시퀀스(sequence of morpheme-tag pairs) 선택하기 절차로 이루어짐. 형태소는 하나 이상의 품사를 가질 수 있으므로 문맥을 고려한 품사 태깅은 형태론적 중의성을 해소하며, 규칙/통계/딥러닝 기반 방법이 있음
a-6. 구문 분석 (Syntax Analysis)	문장 구성 요소들의 문법적 구조를 자동으로 추출하는 기술로, 단어와 절의 계층적 관계를 트리 구조로 표현하는 구구조 구문 분석(Phrase Structure Parsing)과 단어간 계층적인 의존 관계 및 유형을 분석함으로써 문장의 문법적 구조를 의존 트리로 표현하는 의존 구문 분석(Dependency Structure Parsing)이 있고, 각 분석은 규칙/통계/딥러닝 기반 방법으로 수행되며, 기계 번역, 정보 검색, 전문가 시스템 등에서 정확한 문장 의미 분석을 위해 활용
a-7. 의미 분석 (Semantic Analysis)	다의어와 동음어에 의한 어휘적 중의성과 단어 관계 구조에 의한 구조적 중의성을 상황 및 문맥을 반영하여 해소함으로써 문장의 의미를 정확하게 해석하는 기술로, 단어 중의성 해소(Word Sense Disambiguation), 의미역(Semantic Role) 분석(서술어의 의미를 구성 및 보충하는 어휘의 의미적 역할 분석), 의미표현 분석(화자들의 다양한 언어적 표현 파악) 등의 기술 필요
a-8. 문법 교정 (Grammar Correction)	자연어의 문법적인 오류를 감지하고 수정하는 기술
a-9. 키워드 분석 (Keyword Analysis)	문서 집합에 포함된 키워드의 빈도를 조사해 트렌드를 파악하거나 키워드간 연관 관계를 분석하는 기술
a-10. 문서 유사도 분석 (Document Similarity Analysis)	문서간 유사한 정도를 분석하는 기술로, 문서 단어 행렬이나 TF-IDF 행렬 및 코사인 유사도 등을 사용하며, 문서의 복제 여부 확인이나 추천 시스템 구축에 활용
a-11. 텍스트 클러스터링 (Text Clustering)	문서 집합을 구성하는 문서간 유사도를 계산하여 유사한 문서(텍스트)들끼리 몇 개의 군집(cluster)으로 묶어주는 기술
a-12. 정보 검색 (Information Retrieval)	집합적인 정보로부터 원하는 내용과 관련이 있는 부분을 찾아내는 기술로, 사용자 편의를 위해 자동 교정이나 자동 완성 같은 예측 텍스트(predictive text) 기술이나 사용자 언어 습관 등의 행위 패턴에 기반한 추천 기술이 사용되기도 함
a-13. 토픽 모델링 (Topic Modeling)	문서 집합에 내재된 추상적인 주제(토픽) 및 의미 구조를 파악하는 통계적 모델링 기술로, 문서 집합 외에 유전자 정보, 이미지, 네트워크와 같은 자료에 대해서도 적용 가능
a-14. 협업 필터링 (Collaborative Filtering)	많은 사용자들로부터 얻은 기호(taste) 정보와 텍스트 전처리, TF-IDF, 코사인 유사도 등을 사용하여 텍스트 문서로 표현된 콘텐츠(상품 및 서비스) 간 유사도를 계산함으로써 비슷한 콘텐츠를 추천하는 기술

〈표 A-1〉 기타 NLP 기술 영역 (계속)

NLP 기술	설명
a-15. 지식 그래프 (Knowledge Graph)	다양한 개체 및 개체간 관계에 대한 지식을 노드(개체)와 링크(관계)로 표현하는 그래프 모델로, 지식 베이스에 저장된 사실 및 규칙을 그래프로 변환하는 지식 베이스 그래프 모델링과 텍스트 데이터의 단어 및 문장 성분을 추출하여 이들간 관계 지식을 그래프로 변환하는 텍스트 그래프 모델링이 있음. 장점은 개체간 관계를 자유롭게 구조화하여 이질적인 지식 베이스의 통합을 가능하게 하는 스키마(schema) 유연성, 다양한 연관 관계에 기반한 검색 용이성과 새로운 관계 추론 및 풍부한 정보량, 개체간 관계에 대한 우수한 가시성, 네트워크 분석 같은 다양한 그래프 분석 기술 적용 가능성 등이 있으며, 검색, 추천, 질의 응답 시스템에 활용
a-16. 음성 인식 (Speech Recognition)	사람이 말하는 내용을 문자 데이터로 변환하는 STT(Speech-to-Text) 기술로, 음성에 의한 기기제어, 음성 검색, 고유한 음성 패턴에 의한 개인 인증 등에 활용
a-17. 음성 검색 (Voice Search)	사람의 음성 언어를 문자로 변환하여 정보를 검색하는 기술로, 멀티태스킹 환경에 적합한 검색 인터페이스를 제공하고 검색어 입력 속도가 빠른 장점이 있음
a-18. 음성 합성 (Speech Synthesis)	말소리의 음파를 자동으로 생성하는 TTS(Text-to-Speech) 기술로, 사람의 말소리를 녹음하여 일정한 음성 단위로 분할한 다음, 부호를 붙여 합성기에 입력하였다가 필요한 음성 단위만을 다시 합쳐 말소리를 만들어냄. 음성 인식과 함께 번역 기계, 로봇 제조 등에 활용
a-19. 이미지 캡셔닝 (Image Captioning)	주어진 이미지를 묘사하는 텍스트를 생성하는 기술로, 이미지와 텍스트가 결합된 멀티모달 프로그래밍이라고 할 수 있으며, 이미지 캡셔닝을 통한 이미지 검색 성능 제고, 제품 이미지에 대한 설명 자동 생성 등에 활용
a-20. 광문자 인식 (Optical Character Recognition)	손글씨, 타이핑 글씨, 인쇄된 텍스트의 이미지를 기계가 이해할 수 있는 코딩 언어로 변환하는 기술, 또는 이미지에서 텍스트를 추출하는 기술이며 GIF, JPG, PNG, TIFF 이미지 스캔 가능
a-21. 자동 완성 (Auto Complete)	데이터 입력 도중 맨 앞부터 입력한 부분까지의 데이터가 과거에 입력한 것과 동일한 단어나 문장일 경우에 뒤에 이어지는 단어나 문장의 후보가 호출되어 그중에서 선택할 수 있는 기능으로, 긴 문자열을 재입력하는 번거로움을 없애고 입력 오류를 줄일 수 있음
a-22. 데이터 어노테이션 (Data Annotation)	기계학습을 위한 훈련 및 테스트 데이터를 만들기 위해 데이터의 카테고리나 클래스 등 데이터에 대한 설명 정보를 추가하는 주석(annotation) 기술
a-23. 비식별화 (De-identification)	인명 같은 개인 식별자를 삭제 및 마스킹하거나 출생일 같은 준식별자를 누락시키거나 일반화하는 등의 프라이버시 보호를 위한 익명화 기술

◆ About the Authors ◆

**박 현 정** (hyunjpark@korea.ac.kr)

한국과학기술원(KAIST) 경영과학과에서 학사와 석사를 마치고, 서울대학교 경영학과에서 경영정보시스템 전공으로 박사학위를 취득하였으며, 현재 고려대학교 Human-inspired AI 연구소에서 연구교수로 재직 중이다. 주요 연구분야는 빅데이터 분석 및 비즈니스 응용, 자연어처리, AI, 데이터 마이닝, 비즈니스 인텔리전스 등이다.

**임 희 석** (limhseok@korea.ac.kr)

고려대학교 컴퓨터과학과에서 학사, 석사, 박사학위를 취득하였으며, 현재 고려대학교 정보대학 컴퓨터학과 교수로 재직 중이다. 주요 연구분야는 자연어 처리, AI, 브레인-컴퓨터 인터페이스(Brain-Computer Interface), 컴퓨터과학 교육의 인지-뉴로 언어처리(Cognitive-Neuro Language Processing), 정보검색 등이다.