

디지털 격차 해소를 위한 GPT 기반 대화형 키오스크

GPT-based Interactive Kiosks to Bridge

Digital Gaps

채이준
인공지능소프트웨어과
한국폴리텍대학 성남캠퍼스
경기도 성남시, 대한민국
2303110262@office.kopo.ac.kr

하다현
인공지능소프트웨어과
한국폴리텍대학 성남캠퍼스
경기도 성남시, 대한민국
2303110264@office.kopo.ac.kr

조은아
인공지능소프트웨어과
한국폴리텍대학 성남캠퍼스
경기도 성남시, 대한민국
2303110261@office.kopo.ac.kr

이효재¹
인공지능소프트웨어과
한국폴리텍대학 성남캠퍼스
경기도 성남시, 대한민국
lhj00199@kopo.ac.kr

요 약

본 논문은 정보 취약계층의 키오스크 접근성 및 사용자 경험 개선을 목표로, 음성 인식 기술을 활용한 대화형 키오스크를 제안한다. 이를 위해 대화형 주문 시스템, 사용자와의 효과적인 상호작용 기능을 제공한다. 본 연구는 주문 절차의 간소화를 추구하며, 이는 음성으로 주문을 시작하고 알레르기 정보 제공부터 메뉴 및 결제 방법 선택, 최종 결제까지의 주문 전 과정을 포함한다. 이러한 접근이 정보 취약계층의 키오스크 사용 경험을 개선하고 비대면 주문 시스템의 효율성 및 경제성을 높일 것으로 기대한다.

키워드 : 키오스크, 음성 인식, 음성 합성, GPT-4, 정보 취약계층

1. 서론

키오스크(Kiosk)는 ‘정보서비스와 업무의 무인·자동화를 통해 대중들이 쉽게 이용할 수 있도록 공공장소에 설치한 무인 단말기’를 말한다[1]. 키오스크는 인건비, 매장 임대료 등 고정비용 상승 부담뿐만 아니라 COVID-19 팬데믹 시대와 맞물려 언택트(Untact: 비대면) 소비 확산과 함께 빠른 속도의 보급률을 보인다. 키오스크 시장은 2006년 약 600억 원 규모에서 급속도로 성장해 2018년 기준 약 3,000억 원 규모에 달하며, 2023년까지 매년 약 5.7%씩 성장하고 있다[2]. 통계청에 따르면 키오스크의 국내 도입률은 2021년 기준 전체 업종으로는 4.5%, 운영 형태별로는 프랜차이즈 매장이 9.7%이며, 그 중 피자, 햄버거, 샌드위치 및 유사 음식점업은 16.7%에 달하는 만큼 보급률이 높고 보급 속도가 빠르다[3]. 그러나 이러한 시스템은 사용자의

다양성을 충분히 고려하지 않고 설계되었기 때문에 이러한 변화가 오히려 일상의 큰 불편으로 다가오는 사용자층의 디지털 격차가 발생하였다. 노화, 신체적 장애, 언어장벽, 그 밖의 사회·문화적 요소로 인해 스마트기기 등 정보화기기에 대한 접근과 활용에 있어 제약을 경험하는 정보 취약계층[4]의 경우, 기술 발전의 혜택을 동등하게 누리지 못한 채 오히려 부정적인 경험을 얻는 것이다. 노인실태조사 결과보고서에 따르면, 65세 이상의 노인 1만 97명 중 58.1%가 키오스크를 통해 식당에서 주문을 해보았으며, 그중 64.2%가 불편함을 경험했다고 응답했으며[5], 휠체어 이용자가 조작할 수 있는 위치에 작동부가 설치되어 있는 키오스크의 비율은 25.6%에 불과했다. 휠체어에 앉은 사람이 볼 수 있는 곳에 화면이 설치된 키오스크 비율은 36.4%, 음성 정보를 제공하는 키오스크의 비율도 27.8%에 그쳤다[6].

따라서, 본 연구는 정보 취약계층의 키오스크

¹ 교신저자

접근성을 개선하기 위해 음성인식기술을 활용한 대화형 키오스크 개발을 제안하고자 한다.

2. 관련 연구

2.1. 음성 양방향 대화 원리

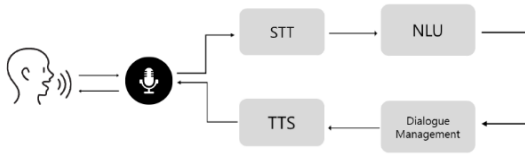


그림 1. 음성 양방향 대화 원리 [1]

자연어 처리 분야 내에서는 텍스트 예측, 분류, 생성, 챗봇, 정보 검색과 문서 랭킹, 지식 기반 엔티티와 관계도, 토픽과 키워드, 텍스트 추론, 텍스트 데이터 상호변환, 가짜 뉴스와 혐오 발언 등이 있다. 여기에 음성 엔진을 연동하면 챗이 아니라 음성을 통해 대화가 가능하고 전화를 연계하면 콜봇이 된다. 음성엔진은 STT(Speech-to-Text)를 이용해 소리를 글자로 바꾸며 NLU(자연어 이해)를 통해 인식, 이것을 Dialogue management(학습 데이터 관리 툴)로 보내어 데이터를 찾고 TTS(Text-to-Speech)에서 글자가 소리로 나가는 원리이다[7].

2.2. 연령대별 키오스크 불편 사례

한국소비자원에서 의뢰한 소비자 설문조사에 따르면 전체 500명 중 233명(46.6%)이 최근 1년간 키오스크 이용 중 불편 또는 피해를 경험했다고 응답했다. 연령대별로 각 유형에 대하여 불편을 느끼는 정도는 차이가 있었으나, 모든 연령대가 ‘뒷사람 눈치’, ‘조작 어려움’, ‘(원하는 상품의) 검색이 어려움’에 대한 불편을 느끼는 것으로 나타났다 [8].

2.3. 배리어프리 키오스크

키오스크의 접근성이 떨어지는 문제는 장애인들에게 또 다른 차별로 다가온다. 때문에 장애인의 키오스크 접근성을 높이기 위해 배리어프리 키오스크의 도입이 필요하다.

배리어프리 키오스크는 휠체어 사용자, 시각 장애인, 청각 장애인 모드가 있어 장애 유형에 상관없이 모두 이용이 가능하다. 키오스크의 높이 조절이 가능하고 음성 안내와 점자 디스플레이가 포함

되어 있다[9]. 그러나, 배리어프리 키오스크의 경우 가격대가 높아서 보편적으로 쓰기에는 무리가 있다.

구분	20대	30대	40대	50대	60대~
1순위	기기 오류 (70.8)	뒷사람 눈치 (52.9)	뒷사람 눈치 (66.7)	뒷사람 눈치 (61.5)	조작 어려움 (53.6)
2순위	뒷사람 눈치 (47.9)	조작 어려움 (51.0)	검색 어려움 (46.2)	조작 어려움 (51.3)	검색 어려움 (42.9)
3순위	조작 어려움 (35.4)	기기 오류 (45.1)	조작 어려움 (41.0)	할인 적용 (38.5)	뒷사람 눈치 (41.1)
4순위	설명 부족 (27.1)	검색 어려움 (35.3)	할인 적용 (38.5)	기기 오류 (35.9)	용어 어려움 (33.9)

그림 2. 연령대별 키오스크 불편 사례[6]

2.4. 자연어 처리 모델: GPT-4

GPT(Generative Pre-trained Transformer) 모델은 Transformer 아키텍처를 기반으로 구축된 신경망 기반 언어 예측 모델로, 프롬프트라는 자연어 쿼리를 분석하고 언어에 대한 이해를 바탕으로 가능한 최상의 응답을 예측한다. GPT는 방대한 언어 데이터셋에서 사전 학습을 수행하여 일반적인 언어 이해 능력을 습득한다. 이후 학습된 데이터를 기반으로 적합한 모델을 생성해내는 파인 튜닝(Fine-Tuning)을 통해 사람과 구두로 대화하는 듯한 자연스러운 문장을 생성해낸다. 또한 어텐션 메커니즘을 사용하여 입력된 텍스트의 다양한 부분 간의 관계를 학습하여 텍스트의 맥락을 더욱 정확하게 이해할 수 있다. GPT 모델의 이러한 특성은 STT 및 TTS 기술에 통합되었을 때, 주문 시스템과 사용자 간의 효과적인 상호작용에 기여할 수 있다[10]. 본 연구는 특히, 이전 GPT-3.5 모델보다 자연어 텍스트 처리 능력이 개선된 GPT-4 모델에 기반을 둔다.

평가 결과의 하단에 적힌 shot은 모델에게 특정 작업을 수행하기 전에 제공되는 예제를 뜻한다. 0-shot의 경우는 모델에게 질문-답변 쌍의 예제를 제공하지 않고 기존에 학습한 지식을 바탕으로 질문에 대한 답변을 생성하도록 요청하는 것을 의미한다. 1 이상의 shot의 경우는 해당 숫자만큼의 질문-답변 쌍을 제공하고 이를 바탕으로 새로운 질문에 대한 답변을 생성하도록 요청하는 것을 의미한다. GPT-4의 벤치마크 결과를 살펴보면, 모든 영역에서 GPT-3.5보다 향상된 결과를 기록하였다. 특히 수학 문제를 처리하는 부분에 있어서 결과가 큰 폭으로 향상된 것을 알 수 있다. OpenAI는 MMLU 벤치마크를 Azure Translate를 사용하여

다양한 언어로 번역하였으며, 스와힐리어, 웨일스어 등의 리소스가 적은 언어들을 포함한 테스트된 언어 대부분이 GPT3.5의 영어 성능보다도 우수하다고 설명하였다[11].

Table. 4 GPT-4 Benchmark Result

	GPT-4	GPT-3.5
MMLU	86.4% 5-shot	70.0% 5-shot
HellaSwag	95.3% 10-shot	85.5% 10-shot
A12 Reasoning Challenge (ARC)	96.3% 25-shot	85.2% 25-shot
WinoGrande	87.5% 5-shot	81.6% 5-shot
HumanEval	67.0% 0-shot	48.1% 0-shot
DROP (F1 score)	80.9 3-shot	64.1 3-shot
GSM-8K	92.0% 5-shot	57.1% 5-shot

그림 3. 자연어 처리모델: GPT-4[11]

3. 제안방법

3.1. 고객 감지 기능

주문을 원하는 고객이 키오스크 기기에 접근할 시, ToF(Time-of-Flight) ranging sensor를 통해 감지하여 고객의 음성 안내 사용 의사를 묻는다. 동시에 화면에는 일반적인 화면 접촉 방식의 주문으로 진입할 버튼이 표시되어 음성 안내를 강제하지 않는다. 고객이 음성 안내 사용에 긍정적인 대답을 하는지 아닌지는 STT(Speech to Text)를 통해 문자로 변환되어 OpenAI에서 제공하는 GPT-4를 이용하여 판별한다. 긍정일 경우에는 주문을 위한 음성 안내가 시작된다. 이후 고객은 순차적인 대화를 통하여 주문할 목록과 결제 수단을 선택하게 된다.

3.2. STT / TTS 기능

대화형 키오스크에서 STT와 TTS 기술의 통합은 사용자와 기계 간의 상호작용을 자연스러운 대화 형태로 전환해준다. 사용자는 복잡한 조작 없이 음성만으로 서비스를 이용할 수 있게 되며, 키오스크는 더욱 다양한 사용자층에 서비스를 제공할 수 있게 된다. 본 논문에서는 Google Cloud의 Speech-to-Text와 Text-to-Speech 기술을 이용한 음성 인식 및 합성을 제안한다. Google Cloud의 STT 및 TTS 서비스는 높은 정확도와 자연스

러운 음성 합성 능력을 제공한다. 이는 구글의 첨단 머신러닝 알고리즘과 대규모 언어 데이터베이스에 기반을 두고 있으며, 다양한 언어와 방언을 지원하여 전 세계적으로 폭넓게 활용될 수 있다.

4. 실험결과

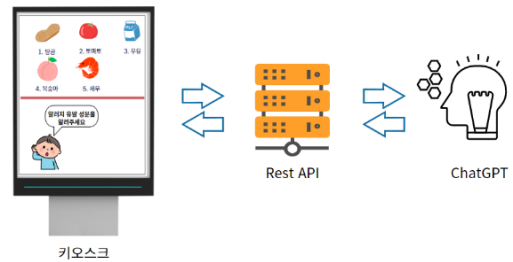


그림 5. 키오스크 구성도

키오스크랑 ChatGPT를 API 형식으로 통신한다. 키오스크에서 사용자 입력을 받은 후, 이를 API요청으로 변환하여 ChatGPT 서버에 전송한다. 이 요청은 사용자의 질문이나 요청 등을 포함 할 수 있다. 받은 요청을 처리하고, 해당 요청에 대한 적절한 응답을 생성한다. 이를 위해 ChatGPT 모델을 사용하여 자연어 처리 및 이해를 수행할 수 있다. ChatGPT 서버는 생성된 응답을 API 응답 형식으로 변환하여 키오스크에 전송한다. 이 응답은 사용자의 질문에 대한 답변이 될 수 있다. 키오스크는 받은 응답을 사용자에게 표시하여 보여준다. 이는 화면에 텍스트로 표시되거나 음성으로 제공될 수 있다.

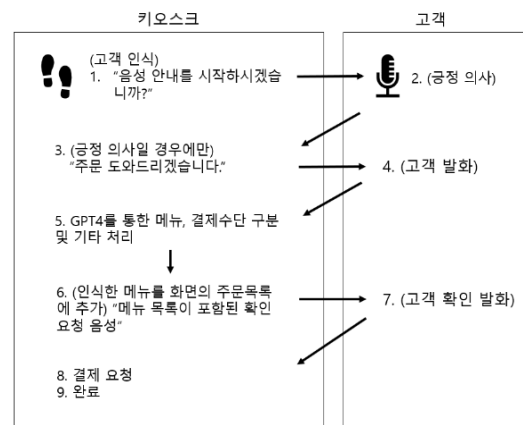


그림 6. 키오스크 흐름

음성 안내 사용 의사를 판단하기 위해 고객을 인식한 후 멘트를 시작한다. 고객이 긍정 의사를

표하면 주문을 시작한다. GPT4를 통한 메뉴, 결제수단 구분 및 기타 처리를 하고 메뉴 목록이 포함된 확인 요청 음성을 표시한다. 고객이 고객의 확인 의사가 오면 결제요청에 마무리가 된다.

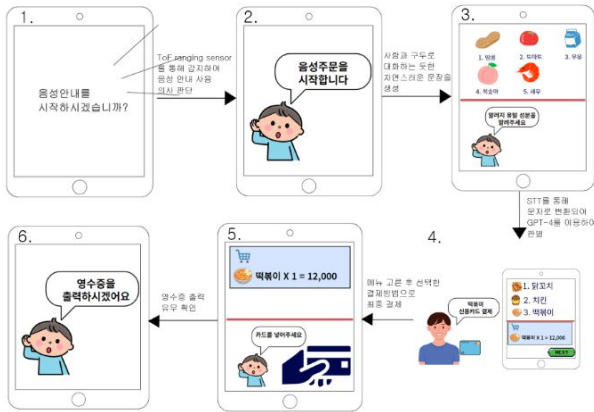


그림 7. 키오스크 흐름

키오스크 흐름 UI이다. 위 그림과 마찬가지로 음성 안내 사용 의사를 판단하면 음성주문이 시작되고 키오스크는 사람과 구두로 대화하는 듯한 자연스러운 문장을 생성하게 된다. STT를 통해 문자로 변환되어 GPT-4를 이용하여 판별한다. 메뉴를 고른 후 선택한 결제방법으로 최종 결제를 하며 영수증 출력유무 확인 후 완료된다.

5. 결과

음성 인식 기반 대화형 키오스크가 정보 취약 계층 등 다양한 사용자층의 키오스크 사용 경험을 개선하고, 다양한 산업 분야에서 운영 효율성을 증대시킬 수 있을 것으로 기대한다. 음성 명령을 통한 조작은 자연스러운 대화를 통해 명령을 내릴 수 있어 기존의 터치 기반 키오스크에 비해 쉽고 빠르며 직관적인 사용성을 선보일 수 있다. 또한 사용자의 점원 의존도를 줄여 운영 효율성 및 비용도 절감할 수 있을 것으로 보인다. 결국, 다양한 사용자층을 포용하여 디지털 격차를 해소함으로써 디지털 대전환기에 혁신적인 해결책이 될 것으로 기대한다.

참고문헌

- [1] 시사경제용어사전
(<https://www.moef.go.kr/sisa/dictionary/detail?idx=>

2602)

- [2] 오정아, 김은희, 서하람. (2023). A Study on Guideline of Unmanned Kiosk Design Based on Universal Design Concept: Focusing on the UI Intuitiveness and Ease of Use of Franchise-Branded Store Kiosks, 2023년
- [3] 통계청. 외식업체경영실태조사. 2022년
- [4] 손지연, 정보취약계층 소비자권익 증진방안 연구. 2017년
- [5] 이윤경, 김세진, 황남희, 임정미, 주보혜, 남궁은하, 이선희, 정경희, 강은나, 김경래. 2020년도 노인실태조사. 2020년도
- [6] 한국정보화진흥원, 2019년 무인정보단말(키오스크) 정보접근성 현황조사**, **. 2019년
- [7] MIT 테크놀로지 리뷰 편집팀, MIT Technology Review Korea - 스페셜리포트
- [8] 한국소비자원, 피해예방정보, 2022년
- [9] 이수정 객원기자, [청년시선] 배리어프리 키오스크의 도입이 필요하다, 2023년
- [10] aws. GPT란 무엇인가요?
- 두산백과. 챗GPT
- [11] OpenAI. GPT-4 Technical Report. 2023년