**2023 제1회 K-디지털 플랫폼 AI 경진대회**

**프로젝트 계획서**

|  |  |
| --- | --- |
| **제 안 명 (사업명/아이템 명)** | 멀티모달을 활용한 시각장애인 보조 애플리케이션 / 뷰파인더 |
| **팀 명** | 엔드포인트 |
| **팀 구성 인원** | 조세은(팀장), 김가람, 양정열, 이승준 |

**1. 개발 배경 및 개발 방향**

|  |
| --- |
| 일상생활 속 전반이 디지털 영역으로 넘어가고 있는 디지털 시대이지만 시각장애인들은 디지털이 주는 이점을 누리기 힘든 현황이다. 프리시던스리서치 등 글로벌 시장조사업체는 2021년 디지털 전환 시장 규모를 4844억 4000만 달러(약 629조 7720억원)로 집계했다.[1] 하지만 2022년 한국소비자원에서 실시한 설문조사에 따르면 쇼핑 애플리케이션과 모바일 애플리케이션 사용 경험이 있는 시각장애인 중 약 92%가 어려움을 겪었다고 응답했다.[2] 이를 통해 디지털 환경에서 접근성 문제가 시각장애인의 일상생활에 큰 영향을 미치고 있다는 것을 알 수 있다.  시각장애인들이 일상생활 속에서 겪는 대표적인 불편함은 시각 정보에 대한 부족한 접근성 및 이동 및 보행의 어려움이다. 청각 또는 촉각 등을 이용한 정보처리 과정의 수행에 크게 의존해야 하는 시각장애인들이 청각, 촉각 등을 이용하지 못하는 상황에 놓인 경우, 이를 해결해 줄 수 있는 방안이 필요하다. 이동 및 보행의 어려움 또한 문제로 지적된다. 2017년 보건복지부가 발표한 장애인 실태조사 결과에 따르면 전체 시각장애인의 83.6%는 혼자 외출한다고 하며, 이들 중 외출 빈도가 거의 매일이라고 답한 시각 장애인은 66.9%로 조사됐다.[3] 이를 바탕으로, 이들을 보조할 수 있는 기기의 필요성이 강조된다.  2020년 말 기준 국내 등록 장애인은 263만 3,000명으로 전체 인구 대비 5.1%를 차지했으며 작년 말 대비 1만 4,000명이 증가했다. 그중 장애 유형 별로 살펴보면 시각장애인은 9.6%를 차지해 장애 유형 중 세 번째로 높은 비율을 보였다. 정부는 장애인의 사회적 참여를 적극 유도하고, 편의를 제공하기 위해 관계 법령 및 시행 지침 수립을 통해 장애인을 위한 편의 시설 설치를 확대하고 보조 기기를 지원하고 있다. 하지만 시각장애인 편의 시설 설치 실태, 보조 기기 지원 현황은 미비한 실정이다. 서울시 4개 센터에서 보유하고 있는 지체 장애인용 보조 기기는 약 1,162개인데 반해, 시각 장애인용 기기는 73개에 불과하다.[4] 이러한 배경 속에서, 본 프로젝트는 멀티모달을 활용해 시각 정보에 대한 부족한 접근성을 보완할 수 있는 애플리케이션의 도입을 제시한다.  본 프로젝트가 개발하고자 하는 애플리케이션에 대하여 설명하겠다. 애플리케이션의 이름인 ‘ViewFinder(뷰 파인더)’ 는 앞이 보이지 않아 일상생활이 힘든 시각장애인들과 색을 구분하지 못해 불편함을 느끼는 색맹인들과 같은 사람들을 위해 시야를 확보해 주겠다는 뜻에서 붙인 이름이다. 본 팀이 개발하려는 ViewFinder의 원리 및 일상 생활에서의 적용 방식은 다음과 같다. 시각적 불편함을 느끼는 사용자가 주변의 시각 정보를 알고 싶을 때, 스마트폰의 애플리케이션 ‘ViewFinder’를 사용한다. 사용자는 애플리케이션을 통해 사진을 찍고, 그 시각 정보에 대해 궁금한 점을 질문한다. 이때, Azure AI 시스템과 연동된 인공지능 모델이 사진과 사용자의 질문을 입력으로 받아 대답을 생성한다. 생성된 대답은 사용자에게 전달되는 방식으로 이 시스템이 구현된다. 자세한 모델 개발 및 애플리케이션 구동은 개발 계획 설명에 서술한다. ViewFinder는 시각적 불편함을 느끼는 사용자가 현재 처한 환경을 인식하고 분석하여 그 환경에 대한 정보를 제공한다. 이는 시각장애인, 색맹인들에게 시각 정보를 제공해 줌으로써 일상적인 환경에서 안전 및 편의성을 높여주며 시각적 불편함을 가진 사람들의 디지털 격차 해소에 기여할 것으로 기대된다.  <참고문헌>  [1] "2030년 시장 규모 2000조원… 디지털전환 격전지로 뜬 한국." 서울신문, 2022. 12. 05, https://www.seoul.co.kr/news/newsView.php?id=20221206017013  [2] 한국소비자원, 장애인 소비자 모바일 거래 실태조사 보고서  [3] 보건복지부, 한국보건사회연구원 2017년 장애인 실태조사 보고서  [4] “서울시 장애인보조기기센터, 시청각장애인용 기기 부족” 에이블뉴스, 2022.11.24, https://www.ablenews.co.kr/news/articleView.html?idxno=100882 |

**2. 개발 계획 설명**

|  |
| --- |
| 본 프로젝트에서 사용할 AI모델과 데이터 애플리케이션 기술 구현에 대해 서술한 뒤 구체적인 개발 계획에 대해서 설명하겠다.  본 팀의 핵심 분석 AI 모델은 멀티모달이다. 멀티모달은 이미지와 자연어를 동시에 처리하는 기술로, 그 중 visual question answering(VQA)에 적합한 모델을 개발한다. 이를 위해, 이미지 처리를 위한 모델과 자연어 처리를 위한 모델을 각각 개발하고, 이들의 특징을 통합 벡터로 결합하는 과정을 거친다. 자연어 처리는 pretrained된 XLMRobertaModel을 이용하고, 이미지 처리는 pretrained된 ResNet50을 이용한다. 다언어 텍스트의 대규모 데이터셋을 사용하여 사전 훈련된 XLM-RoBERTa는 한 언어에서 학습한 지식을 다른 언어에도 적용할 수 있다. 이는 학습 데이터가 부족한 언어에 대해서도 효과적인 성능을 발휘 수 있어 모델로 채택한다. 이미지 처리는 사전 훈련된 ResNet50을 사용하며, 최종 fully connected layer의 출력 노드 수를 BERT의 출력인 768과 동일하게 설정한다.이를 통해 이미지와 자연어 각각에서 동일한 크기의 768 벡터 특징값을 추출하고, 이들을 point-wise 연산을 통해 통합 벡터로 결합한다. 이 통합 벡터를 사용하여, 주어진 질문에 대한 답변 후보 목록 중 가장 높은 확률 값을 가진 출력을 선택하는 답변 모델을 구현한다.  본 팀이 모델 구현에서 활용할 데이터 셋은 다음과 같다. AI hub에서 제공하는 한국어 visual question answering 데이터 셋을 활용한다. https://www.aihub.or.kr/  aihubdata/data/view.do?currMenu=&topMenu=&aihubDataSe=data&dataSetSn=71357 에서 확인할 수 있으며 공공 데이터 셋으로 수집이 가능하다. 본 데이터 셋은 인간이 가진 상식이나 배경 지식을 바탕으로, 이미지에 관련한 질문에 대해 이미지 속에서 답을 찾아야 하는 task를 위해 구축된 데이터 셋이다. 본 팀의 task는 특별한 상황에 대응하는 것이 아닌, 일상 생활에서 생겨나는 간단한 질문에 답변을 제공하는 것이라 해당 데이터 셋이 task에 잘 부합한다고 판단한다.  본 애플리케이션 기술 구현의 구체적인 방법은 다음과 같다. AI 모델은 로컬에 저장하고 서버는 파이썬 웹 프레임워크 Flask를 사용하여 구축한다. 구축한 Flask 서버의 API를 사용하여 안드로이드 스튜디오와 연동한다. 안드로이드 스튜디오에서 네트워크 요청 처리과정은 다음과 같다. 먼저 Manifest 파일에 네트워크 권한을 추가하여 애플리케이션이 인터넷에 액세스 할 수 있도록 허용하고 Retrofit과 OkHttp를 함께 사용하여 서버에 데이터를 전송하고 응답을 처리하는 과정을 구현한다. 사용자는 질문을 직접 텍스트로 입력할 수도 있고 음성 인식으로 입력할 수도 있다. 음성 인식의 경우 안드로이드 내장 API로 지원하고 있는 Google SpeechToText 기술을 사용하여 음성을 텍스트로 변환하는 기능을 사용한다. 이를 바탕으로 사용자가 촬영한 사진과 입력한 질문에 대한 RequestBody를 작성하고 해당 데이터를 서버에 요청하고 응답을 처리한다. 응답은 JSON 파일로 반환하고 결과값을 텍스트로 받아온다. 받아온 텍스트는 안드로이드에서 내장 API로 지원하고 있는 Google TextToSpeech 기술을 사용하여 결과를 음성으로 변환하여 처리한다. 이렇게 구성된 애플리케이션을 구현한 모델과의 통신을 통해 사용자에게 음성적 피드백을 제공할 것이다.  위 이론들을 기반으로 한 개발 계획은 다음과 같다. 크게 애플리케이션 환경 구축과 모델 구축 2가지에 대해서 준비를 할 예정이다. 안드로이드 애플리케이션 및 서버 환경 구축에 대해 설명하겠다. Flask를 활용하여 간단한 서버를 구축한 뒤 API를 안드로이드 스튜디오와 연동하여 서버와 통신할 계획이다. 모델 구축은 Colab에서 이미 한 번 시험해 본 바 있으므로, 생성한 모델을 로컬에 저장한뒤 flask 서버 내에서 실행될 수 있도록 구현할 예정이다. 이를 바탕으로 실제 모델을 안드로이드 애플리케이션에 배포해 시범적으로 테스트하며, TTS(Text To Speech) 및 STT(Speech To Text) 기능의 구현을 목표로 하고 있다. 본 프로젝트에서 사용할 AI모델과 데이터 셋 그리고 Azure를 적용한 애플리케이션 기술 구현에 대해 서술한 뒤 구체적인 개발 계획에 대해서 설명하겠다. |

**3. 시제품(결과물) 형태 및 활용 방안**

|  |
| --- |
| 본 프로젝트의 시제품 ‘ViewFinder’는 안드로이드 애플리케이션으로 시각 장애인과 색맹인 사용자들이 주변 환경을 더 잘 이해하고 탐색할 수 있도록 설계된다. 시각 장애인과 색맹인을 위한 결과물로써 사용자가 일상 생활에서 마주하는 다양한 상황들을 AI 기술을 활용한 솔루션을 바탕으로 보다 쉽게 해결하고 디지털에 대한 접근성을 향상 시킬 수 있도록 돕는 것이 핵심이다. 아래 그림은 본 애플리케이션의 활용 방안을 나타낸 2가지 상황이다. 먼저 시각장애인이 버스를 타기 위해 버스 정류장을 찾고 있는 상황이다. 해당 장소가 버스 정류장이 맞는지 확인하고 싶을 때 본 애플리케이션을 활용해 도움을 받을 수 있다. 확인하고 싶은 장소의 사진을 촬영하고 “앞에 보이는 곳이 버스 정류장 맞아?” 라는 질문을 입력하면 애플리케이션은 이미지에 관련된 질문에 대해 이미지에서 답을 찾는 task를 진행하게 된다. 해당 task를 처리하고 난 결과를 사용자에게 전달해 사용자가 보다 빠르고 쉽게 주변 환경을 인지할 수 있도록 돕는다. 두 번째 상황은 색맹인들이 횡단보도에서 신호등의 교통 신호를 기다리고 있는 중이다. 색맹은 적색과 녹색을 구분하지 못하는 ‘적록색맹’이 가장 많은데 신호등의 교통 신호를 잘못 보는 경우가 있을 수 있다. 이럴 때 사용자가 신호등을 촬영하고 “지금 신호등이 무슨 색이야?” 라고 질문을 입력하면 해당 task의 처리 결과 값으로 무슨 색인지 전달 받게 된다. 전달 받은 결과를 바탕으로 색맹인들은 보다 안전하게 생활하고 교통을 준수할 수 있다. 또한 일생 생활에서 색을 구분하지 못하는 여러 상황에 직면했을 때 본 애플리케이션을 활용하여 불편함을 해소할 수 있다. |

**4. 기대효과 및 오픈(상용화) 계획**

|  |
| --- |
| 본 프로젝트는 다음과 같은 기대효과를 가진다.  첫째, 디지털 혁신의 사각지대에 있는 사람들에게 배리어프리 환경을 제공한다. 배리어프리(Barrier-free)는 장애인 및 노인과 같은 사회적 약자들이 편하게 살아갈 수 있게 물리적인 장애물, 심리적인 벽 등을 제거하자는 운동 및 정책을 말한다. 디지털 환경에서의 배리어프리는 사용자가 쉽게 이용할 수 있도록 설계된 것을 의미한다. 시각, 청각 등 다양한 장애를 가진 사람들도 디지털 기술을 활용하여 정보에 접근하고 소통할 수 있어야 한다. 이를 기반으로 본 프로젝트는 디지털 혁신의 사각지대에 놓인 시각장애인들을 위한 ‘ViewFinder’ 애플리케이션을 개발하여 사회적 약자들이 편하게 살아갈 수 있도록 도와준다.  둘째, 시각장애인들의 이동권을 보장한다. 시각 장애인들은 주로 흰 지팡이인 '케인'(Cane)을 이용해 점자 블록을 읽으며 걷는다. 그러나 이는 여전히 주변 환경의 한정된 정보만을 제공하며, 특히 복잡한 도심 환경 속에서 이동은 여러 제약이 따른다. 'ViewFinder'는 이러한 제약을 극복한다. 사진에 대한 구체적인 설명을 제공함으로써 시각장애인들이 주변 환경을 더 정확하게 인식하고 이해할 수 있게 도와준다. 이는 시각 장애인들의 불편함을 해소하며 그들의 이동권 확장까지 기여할 수 있다.  셋째, 복지 산업 및 기타 산업으로의 기술 확장이 가능하다. ’ViewFinder’ 에서 활용한 기술은 시각 장애인과 색맹인에게 도움을 줄 수 있는 보조 장치로서 복지 산업에 활용될 수 있다. 또한, 도로를 따라 주행하며 교통 표지판을 보고 상황에 맞는 판단이 가능한 무인 자동차, 사용자만의 개인 비서 역할을 수행하는 애플리케이션로써의 기술 확장이 가능하다. 멀티모달을 활용한 Visual Question answering(VQA) 기술을 산업 전반에 확대함으로써 다양한 산업의 활성화를 기대할 수 있다.  이를 바탕으로 멀티모달 기반의 AI모델 디지털 기술은 시각적 불편함을 느끼는 사람들에게 시각 정보를 제공하고 현대 사회의 중요한 과제 중 하나인 디지털 격차 해소에 기여한다.  장애인을 위한 보조기기 시장은 세계적인 고령화 추세와 후천적 장애 발생 증가에 따라 매년 큰 폭으로 성장하고 있다. 2017년 Coherent Market Insights Analysis에서 발표한 전 세계 보조기기 시장 규모는 2015년 140억 달러이며, 2024년까지 약 7.4% CAGR(연평균 성장률) 수준으로 지속 성장할 것을 예측하고 있다. 국내 재활 및 보조기기 시장 규모 또한 2018년 1조 3,726억 원에서 연평균 4.1%성장률로 증가하여 2024년 약 1조 7,468억 달러 규모로 성장할 것으로 전망된다.[1] IT기술의 발달로 보조기기 시장의 방향 역시 달라졌다. 단순 보조 용구에서 생활 지원기기 및 애플리케이션 등 장애 보조 시스템 전반의 영역이 크게 발달될 것으로 예상된다. 하지만 지속적으로 성장하는 보조기기 시장임에도 불구하고 국내 시각장애인 보조 애플리케이션의 현황은 미흡한 실정이다. 기존에 존재하는 시각장애인 보조 애플리케이션의 경우 정해진 형식 내에서 정보를 가져올 수 있지만 원하는 특정 정보를 제공하지는 않는다. 이러한 기존의 애플리케이션과는 달리 ‘ViewFinder’ 는 사용자가 원하는 정보에 대해 적절한 답변을 제공한다. 이를 통해 시각장애인들의 수요를 충족시키고 지속적으로 안정적인 서비스를 제공함으로써 시장에서 높은 시장점유율을 기대할 수 있다.  현 애플리케이션을 상용화 시, 사용자들의 피드백을 지속적으로 수집하고 피드백을 바탕으로 새로운 기능 추가 및 버그 수정을 통해 애플리케이션을 개선한다. 또한 시각적 불편함을 가진 사용자들을 대상으로 한 애플리케이션이므로 사용자들이 더 편리하게 활용할 수 있도록 사용자 경험을 고려하여 UI/UX를 디자인 한다. 본 애플리케이션은 안드로이드 플랫폼을 기반으로 개발되어 향후 더 넓은 사용자에게 서비스를 제공하기 위해 iOS 플랫폼으로의 확장성을 기대할 수 있다. 또한 모바일 애플리케이션 뿐 아닌 웹 서비스로의 확장을 통해 브라우저 상에서 서비스를 제공할 수 있다. 이러한 플랫폼 확장을 바탕으로 사용자의 접근성과 편의성을 높여 디지털 격차 해소에 기여할 수 있다.  <참고문헌>  [1] 2018-2020 중소기업 기술로드맵(의료서비스·기기) |