

정신건강 분야를 위한 공감형 AI 에이전트 기술 동향

가천대학교 | 고효진·유소엽·정옥란

1. 서론

코로나19로 인한 팬데믹을 겪으면서 사회는 빠른 속도로 변화했고, 그 안에서 우울, 고립, 불안 등 다양한 정신건강 문제에 직면하게 됐다. 정신건강 관련 문제들이 기존에는 개인의 문제로만 여겨졌다면, 최근에는 사회적인 문제로 판단되며 중요한 이슈로 떠오르고 있다. 건강보험심사평가원의 국민 관심 질병 통계에 따르면 우울증 환자 수가 2017년 680,169 명에서 2021년 910,785 명으로 34% 가량 증가하였다[1]. 2023년 12월 13일 정부의 고립·은둔 청년 실태조사 결과 발표에 따르면 국내 고립·은둔 청년이 약 54만 명에 달하고, 특히 고립이나 은둔을 경험한 청년의 75.4%가 자살을 생각했던 것으로 나타났다[2].

이처럼 정신건강 질환으로 어려움을 겪는 사람이 늘어났음에도 불구하고 우울증을 비롯한 정신 질환은 의지력이나 정신력의 부족으로 인한 증상이라 생각하는 편견으로 인해 정확한 원인 파악이나 치료를 받지 않는 경우가 많다. 국립정신건강센터 정신건강연구소에 따르면 2022년 기준 지난 1년간 정신건강 문제 경험자가 63.9%임에도 불구하고 이들 중 26.9%만 전문 인력 외에 가족, 친구 등 타인에게 도움을 요청한 것으로 나타났다[3]. 또한 주변의 부정적 시선, 정신질환은 스스로 극복해야 한다는 인식 등을 이유로 병원 치료를 고민한다고 응답했다.

사회적으로 정신건강에 대한 경각심이 높아졌음에도 편견으로 인한 낮은 접근성, 전문 인력 부족 등의 한계를 해결하기 위해 최근 정신 의학 분야에서는 인공지능(AI) 등의 기술을 적극적으로 활용하는 연구가 증가하고 있다. 특히 많은 사람들의 정신건강 분야에 대한 접근성을 높이고 간단한 상담을 제공할 수 있는 챗봇에 대한 연구와 서비스가 다양하게 존재한다.

챗봇은 인공지능과 자연어 처리 기술을 이용해 인간 수준의 대화를 시뮬레이션 하는 AI 기반 소프트웨어이며, 대표적인 에이전트 기술이다. 최근 ChatGPT[4],

Bard[5]와 같은 생성형 AI 챗봇들이 공개되고 높은 성능을 보여주면서 일상 속에서 AI 챗봇의 범용적인 활용 가능성이 매우 높아지고 있다. 최근에는 법학, 의학 등 다양한 분야에 접목하기 위한 연구가 활발하게 진행되고 있다.

하지만 의학, 특히 정신건강 분야에서는 범용적인 AI 챗봇을 바로 적용하기에는 여러 한계점이 존재한다. 정신건강 분야에서 AI 챗봇이 적극적으로 활용되기 위해서는 의학적인 전문 지식을 탑재하고 있어야 한다. 또한 단순히 작업 처리를 위한 수준의 자연어 이해 능력이 아닌 감정, 의도, 맥락 등의 비언어적이고 추상적인 정보를 이해하고, 이를 기반으로 대화를 생성할 수 있어야 한다.

전문적인 지식을 갖춘 공감 기반의 AI 에이전트를 정신건강 분야에서 활용하게 되면 대화를 통해 정신건강 질환의 분류, 예측 등의 전문적인 기능을 수행할 수 있을 뿐만 아니라 위로와 공감 등을 통한 치유의 기능도 수행이 가능하다. 본 원고에서는 이러한 정신건강 분야의 특성을 반영한 공감형 AI 에이전트 기술에 대해 살펴보려고 한다.

본 원고의 2장에서는 정신건강 분야를 위한 공감형 AI 에이전트를 구축하기 위한 주요 기술을 소개하며, 3장에서는 국내외 기술 동향을 다양한 사례와 함께 정리 및 소개한다. 4장에서는 정신건강 분야에서 공감형 AI 에이전트를 활용할 수 있는 방안을 제안하고 마지막으로 5장에서는 정신건강 분야에서 AI서비스 활용을 강화하기 위한 방안과 함께 결론을 맺는다.

2. 정신건강 분야를 위한 공감형 AI 에이전트 주요 기술

정신건강 분야를 위한 공감형 AI 에이전트는 기존의 AI 챗봇과는 달리 추가적으로 필요한 기술들이 존재한다. 정신건강을 다루는 분야적인 특성상 감정적으로 사람을 이해하고 공감할 수 있어야 할 뿐만 아

나라 전문적인 의료 지식을 기반으로 하고 있어야 한다. 본 장에서는 정신건강 분야를 위한 공감형 AI 에이전트를 구축하기 위해 필요한 필수적인 주요 기술들을 소개한다.

2.1 멀티모달 기반 감정 인지·분석

공감형 AI 에이전트는 사용자와 상호작용함과 동시에 사용자의 감정을 인식, 분석해야 하며 이에 그치지 않고 에이전트 역시 공감적 행위를 수행해야 한다. 일반적인 공감형 AI 에이전트 설계를 위한 감정 인지 처리 과정은 감정 자극 평가, 감정 인식, 행동의 감정 조절의 세 단계로 구성된다 [6]

감정 자극 평가는 에이전트가 환경에 대한 감정 평가를 수행하는 것으로써 에이전트가 지각하는 모든 자극이 특정 감정값과 연관되어 에이전트의 정의적 상태를 업데이트하고, 전반적인 행동을 정의하는 과정이다.

감정 인식 단계에서는 에이전트의 감정 상태를 파악하는 것을 목표로 한다. 감정 평가 결과에 기초해 감정 상태를 인식하고, 인식된 사용자 감정으로 정의적 상태를 업데이트한다.

마지막 행동의 감정 조절의 단계에서는 사용자 인지 처리에 대한 에이전트의 감정 조절의 정도를 결정할 수 있다. 앞서 업데이트된 AI 에이전트의 감정 상태로 에이전트 행동의 기본이 되는 인지 처리를 조절하는데, 특히 지각, 주의 집중, 의사 결정과 같은 행동에 감정적인 영향을 준다. 업데이트된 특정 유형의 감정 및 상황과 일치시켜 에이전트의 언어적, 비언어적 행동을 결정할 수 있다.

공감형 AI 에이전트는 사용자의 감정 자극을 평가하고, 이를 기반으로 에이전트의 감정 상태를 업데이트 한 후, 에이전트의 행동을 결정하게 된다. 사용자의 감정은 발화, 표정, 음성의 높낮이, 행동, 심박수 등 다양한 요소들이 복합되어 표현된다. 따라서 공감형 AI 에이전트는 텍스트 뿐만 아니라 다양한 형태의 정보를 복합적으로 이해하고 판단할 수 있어야 하기 때문에 멀티모달 기반 감정 인지·분석 기술이 필요하다.

그림 1은 멀티모달 기반 감정 인지의 과정 및 정보 추출을 보여준다. 사용자의 다양한 행동 데이터를 수집하고 전처리하여 멀티모달 데이터로 구성한다. 입력되는 데이터의 타입에 따라 데이터를 처리하는 기법이 다양하기 때문에 타입별 특징을 추출 후 추출된 특징들을 혼합해 감정, 행동 패턴 등 원하는 정보를 얻는다. 딥러닝 모델을 사용해 다양한 유형의 데이터를

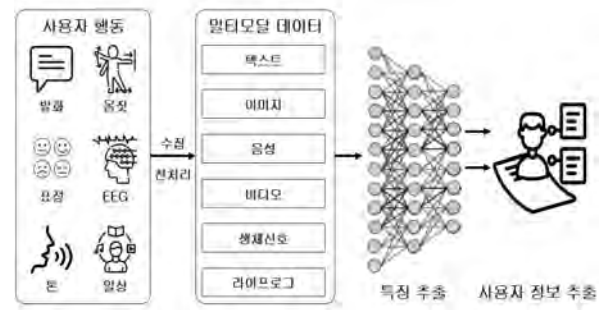


그림 1 멀티모달 기반 사용자 정보 추출

학습할 수 있다. 특히 이미지·영상 데이터 처리를 위한 합성곱 신경망 모델인 CNN[7], 순차 데이터나 시계열 데이터를 처리하는 인공 신경망 모델인 LSTM[8]을 사용해 특징을 추출한다. 또한 Transformer[9] 기반 대규모 언어 모델인 BERT[10]를 사용하여 특징 추출, 혼합 후 원하는 정보를 얻거나 예측할 수 있다 [11][12].

2.2 공감형 대화 생성

인간과 비슷한 수준의 대화를 생성을 위해 언어모델이 사용된다. 언어 모델은 주어진 문장에 대해 임베딩 표현을 생성한다. 대표적인 생성형 언어 모델인 GPT[13][14]가 많이 사용된다. 기존 딥러닝은 라벨링된 데이터가 필요하지만, GPT는 Generative Pre-training Transformer의 약자로, 딥러닝, Transformer[8], Few-shot learning[15], 자기지도학습을 기반으로 한 준지도학습적 접근을 사용하기 때문에 라벨링된 데이터 없이도 다양한 작업에 적용할 수 있는 범용 표현을 학습한다. 대규모 데이터의 패턴과 특성을 파악하여 생성한 합성 데이터를 샘플로 스스로 학습한다. 학습 시에는 대량의 데이터를 처리하기 위해 CNN[7]이나 GAN[16]을 사용한 딥러닝 기법으로 학습을 진행하고 Transformer[9]의 디코더 블록만을 사용해 순차 데이터 내 개별 데이터 관계를 추적해 맥락과 의미를 학습할 수 있게 되었다. 하지만 자원이 적은 도메인에서도 효과적으로 작용해야하기 때문에 저자원 환경에서도 학습이 가능한 Few-shot learning[15]을 적용하였다. few-shot learning으로 학습을 진행하면 소량의 샘플만으로도 작업을 수행할 수 있기 때문에 사람이 직접 라벨링하는 과정이 필요 없이 AI가 스스로 라벨링하는 자기 지도 학습이 가능해진다.

단순 대화 생성 기법이 아닌 공감형 대화를 생성해야 하므로 공감적 표현이 많이 드러난 데이터 세트를 대상으로 모델을 학습시킨다. 특히 정신 건강 분야에 공감형 AI 에이전트를 적용하기 위해서는 AI 허브에

서 제공하는 웰니스 대화 스크립트 데이터 세트[17]와 같이 정신 건강 상담 기록을 가공해 제작한 데이터 세트를 대상으로 사전학습을 한 챗봇은 공감형 대화 생성이 가능하다.

2.3 지식 기반 대화

기존의 대화형 AI 모델이 단발성 대화에 의존하는 문제를 해결하기 위해 대화를 모두 활용해 추출된 정보를 언제든지 활용하여 답변을 생성할 수 있도록 하는 연속 학습 지식 기반 챗봇이 등장하였다. 지식 기반 챗봇은 사용자의 발언에서 직접 추출한 데이터를 지식 기반으로 추론하거나, 질문, 관찰을 통해 정보를 습득한다[18]. 그러나 답변 추론 시 단편적인 정보만을 활용하는 기존의 지식 기반 챗봇은 사회적, 문화적으로 자연스럽게 학습된 상식을 사용하는 인간과 같은 수준으로 대화를 생성하는데 한계가 있다. 또한 기존의 지식 그래프만으로 정보를 추론한다면 새로운 지식을 사용자와의 대화나 행동만으로 얻어야 하므로 지속적인 지식 습득과 개인화 학습에 한계가 있다.

이러한 문제를 해결하기 위해[19][20]에서는 자동 확장형 지식 그래프인 PolarisX[21]를 활용한 연속 학습 기반 대화형 AI 에이전트를 제안한다. 실시간으로 수집된 데이터를 사전학습 언어 모델인 BERT[10]를 사용해 분석 후 관계를 추출한다. 이후 단어와 단어를 관계로 연결한 지식 그래프와 연결함으로써 지능형 챗봇을 구축한다. 지식 기반이 아닌 자동 확장 지식 그래프를 활용한다면 시간의 흐름을 반영해 새롭게 축적되는 대화를 손실이 되지 않은 기존 정보와 함께 연속적으로 학습하는 것이 가능하다.

3. 정신건강 분야 공감형 AI 에이전트 관련 기술 동향

본 원고에서는 정신건강 분야의 공감형 AI 에이전트와 정신건강 분야 공감형 AI 에이전트의 기반이 되는 대화형 AI, 지식 기반 챗봇, 감정형 챗봇에 대해 소개한다. 대화형 AI부터 공감형 AI가 정신건강 분야 적용되기까지 제공된 대표적인 서비스들을 표 1에서 비교한다.

표 1은 AI 기술을 활용한 대표적인 챗봇들을 분류하고, 특징과 활용한 모달리티를 정리해서 보여준다. 챗봇을 분류하는 기준은 매우 다양하지만 본 원고에서는 크게 일반적으로 모든 분야에서 활용하는 범용적인 목적의 챗봇과 정신건강 분야를 목적으로 하는 챗봇으로 분류한다. 또한 대화 생성이 규칙 기반인지, 생성형 모델을 기반으로 하고 있는지, 그리고 공감기술을 적용했는지를 중심으로 각 모델이 해당 특징을 포함하면 ○로 표시한다. ○*로 표현한 모델은 일상 대화에서는 생성모델을 일부 활용하지만, 대화 중 특정 키워드 혹은 지정된 발화가 나타나면 해당 키워드에 대해 사전 지정된 대답 셋에서 검색(retrieval)을 통해 답변하는 모델들을 의미한다. 활용 모달리티는 각 모델들이 대화를 주고받기 위해 활용하고 있는 데이터의 형태를 텍스트, 이미지, 음성, 비디오, 생체신호, 라이프로그로 분류하여 보여준다.

3.1 대화형 AI 챗봇

대화형 AI 챗봇은 인간과 자연어를 사용해 대화함으로써 원하는 작업을 수행한다[22]. 특히 기계학습과 딥러닝의 비약적인 발전으로 인해 대화형 AI 에이전

표 1 AI 기술을 활용한 챗봇 비교표

분류	모델	특징			활용 모달리티					
		규칙 기반	생성모델 기반	공감기술 적용	텍스트	이미지	음성	비디오	생체신호	라이프로그
범용 AI 챗봇	ALICE[32]	○			○					
	Emma[36]			○	○				○	○
	GPT-4[28]		○		○	○	○	○		
	Bard[5]		○		○	○	○	○		
	Gemini[29]		○		○	○	○	○		
정신건강 분야 AI 챗봇	ELIZA[24]	○			○					
	PARRY[25]	○		○	○					
	Wysa[40]	○*			○					
	Tess[41]	○*		○	○					
	Woebot[43]	○*			○					
	Replika[44]	○*	○	○	○	○				○

트는 강화 학습을 통해 스스로 지속 학습을 하고, 사용자 상호작용을 개선해 기존의 기계학습 기반 AI보다 한층 더 높은 수준의 대화가 가능하다. 대화형 AI는 자연어 형태의 입력으로부터 도메인 분류와 엔티티 추출을 통해 자연어 입력을 이해한다. 이를 통해 사용자의 의도를 파악하고 추출된 정보를 기반으로 적절한 답변을 생성할 수 있다[23].

초기의 대화형 AI ELIZA[24]는 사람이 컴퓨터에 입력한 문장에 패턴 매칭과 치환법을 적용해 후보 응답 문장 목록의 쌍을 이루어 작동하는 규칙 기반 모델이다. 또 다른 규칙 기반 모델인 PARRY[25]는 조현병 환자를 흉내 내 입력된 언어의 가중치를 변경해 가며 응답을 생성하는 모델이다. 그러나 규칙 기반 모델은 사람이 직접 생성한 규칙에 따르기 때문에 사용자의 입력 오류에 대응하기가 어려우며 마지막 입력 메시지만 고려하는 단방향 대화가 아닌 경우 잘못된 응답을 생성하는 경우가 많다.

검색 기반 대화형 AI는 API를 사용하여 사용할 수 있는 리소스를 쿼리하고 분석하기 때문에 규칙 기반 모델보다 더 유연한 대화를 제공할 수 있다. 검색 기반 모델은 답변 매칭을 적용하기 전에 인덱스에서 일부 답변 후보를 검색한다[26]. 사람과 사람 또는 대화형 AI의 기존 대화를 사용해 발화를 생성하는 데이터 기반 모델[27]도 연구되었으나 이와 같은 대화형 모델들은 대화가 길어지면서 이전 정보를 기억하지 못한다는 문제가 있다.

3.2 멀티모달 AI 챗봇

멀티모달이란 텍스트, 이미지, 영상, 음성 등 다양한 종류의 데이터를 통틀어 말한다. 일상생활에는 텍스트뿐만이 아니라 여러 가지 데이터가 복합적으로 존재한다. 인간은 일상생활에서 얻는 복합적인 정보들로부터 개념을 통합하여 문맥을 이해하고, 새로운 정보를 습득·추론한다. AI가 인간과 같이 다양한 정보를 받아들여 공유, 소통, 사고, 공감하는 대화를 하기 위해서는 인간과 같은 방식으로 데이터를 다룰 수 있어야 하므로 멀티모달 AI에 대한 관심이 높아지는 추세이다.

초기의 멀티모달 AI는 언어 모델에 시각 정보를 접목하는 것으로부터 시작되었다. OpenAI의 GPT-4[28]는 기존의 텍스트만 입력이 가능했던 생성형 언어 모델 GPT에서 더 나아가 이미지, 영상, 음성까지도 입력할 수 있도록 설계되었다. 사용자가 입력하면 GPT-4는 이미지에 무엇이 표현되었는지, 어떤 이벤트가 발생 가능한지까지 정보 추출부터 예측과 추론까지의

태스크를 수행할 수 있다. 이미지뿐만 아니라 음성을 통한 상호작용, 영상 처리까지 멀티모달 데이터를 한번에 처리하면서 주변 환경에 대한 이해를 기반으로 작업을 수행할 수 있다.

비슷한 멀티모달 AI 챗봇으로는 구글의 Bard[5]가 있다. Bard는 생성형 AI 어시스턴트로, GPT-4와 마찬가지로 텍스트, 음성, 사진 형태의 대화와 명령 처리가 가능하다. 또한 구글의 서치 엔진을 사용하기 때문에 GPT-4를 능가하는 실시간 최신 데이터 기반 답변 제공이 가능하고 출처가 불명확한 데이터들을 조합하여 답변을 생성하는 GPT와 다르게 답변의 출처까지 제공 가능하다.

최근 구글은 Bard에 최신의 멀티모달 AI 모델인 Gemini[29]를 도입하겠다고 발표하였다. Gemini는 멀티모달 데이터로부터 복잡한 추론까지 가능한 멀티모달 AI 언어 모델이다. Gemini는 각 양상의 데이터를 따로 처리 후 통합하여 정보를 추출하였던 이전의 멀티모달 AI와 달리 Cross modal attention network[30]를 사용해 동시에 입력 데이터를 처리하면서 다른 양상 간의 관계와 종속성을 학습한다. 또한 cross modal attention network에서 생성된 풍부한 표현을 통해 이미지 캡션 생성, 텍스트-이미지 생성, 코드 생성과 같은 복잡한 작업 역시 수행할 수 있다.

따라서 이러한 멀티모달 AI를 통해 인간과 가까운 방식으로 다양한 정보를 복합적으로 받아들일 수 있으며, 이는 환경, 신체 상태, 생활 방식 등 다양한 요인으로부터 발병하는 정신 질환의 원인을 파악하고 치료법을 제공할 수 있도록 하는 정신 건강 AI 챗봇에 필수적이다. 하지만 현재까지 제안된 멀티모달 AI 에이전트들은 시청각적 데이터만을 다루고 있기 때문에 정신건강 분야에 적용하기에는 EEG 데이터, 건강 데이터, 상담·진료 기록 등 의료 데이터를 복합적으로 다루는 데에는 한계가 있다. 또한 방대하고 비전문적인 데이터를 활용하기 때문에 의학적인 태스크를 수행하는 데에도 한계가 있다. 따라서 CBT[31] 기반 대화, 전문 의료인과의 연결과 같은 기능을 추가하여 정신건강을 위한 멀티모달 AI 챗봇이 필요하다.

3.3 공감형 AI 챗봇

처음 제안된 규칙을 기반 공감형 챗봇인 PARRY[25]에 이어 패턴 매칭 기법을 재귀적으로 시행하는 또 다른 규칙 기반 공감형 챗봇 ALICE[32]가 제안되었다. 그러나 앞서 언급한 규칙 기반 모델의 한계로 인해 최근엔 딥러닝과 인공지능을 사용해 공감형 챗봇을 구축한다[33]. 또한 감정 임베딩 또는 모델링을

강화 학습에 도입한 연구[34][35]는 인코더-디코더 구조로 되어 있는 Seq2Seq를 사용한다.

BERT, GPT와 같은 대규모 언어 모델 등장 이후 감정형 챗봇 역시 새로운 국면을 맞이하였다. Emma[36]는 사용자가 대화를 시작하면 사용자의 감정 클래스를 감지하고 원인을 인식하는 온라인 공감 챗봇이다. 만약 감정 원인이 감지되면 Emma는 실제 심리학자의 상담 전략을 사용해 사용자의 추가 참여를 유도한다. 그렇게 감지된 감정 클래스와 원인을 토대로 공감적인 응답을 생성한다. 이때 GPT를 기반으로 다중 multi-head self-attention mechanism 블록으로 구성되어 있는 Emma[36]는 사용자 질의와 감정 정보의 임베딩을 하나로 통합하여 답변을 생성한다.

따라서 최근 공감형 챗봇은 사용자의 요구에 맞게 개인화된 응답을 생성할 수 있기 때문에 원인과 치료법이 매우 다양한 정신 질환 분야에서 유용하게 사용할 수 있다.

3.4 정신건강 분야 AI 챗봇

스마트기기의 등장으로 실시간으로 생활 데이터를 수집하고 분석하는 것이 가능해졌다. 일상생활에서 얻어지는 개인의 맥락 및 행동 정보를 입력으로 심리, 정신 건강 상태를 예측하는 것 역시 가능하다. 이에 따라 개인의 일상 데이터와 PHQ-9[37] 우울 설문, 대화 데이터를 종합하여 상담과 진단을 제공하는 인공지능 가상 치료사 연구에 대한 관심도 역시 높아지고 있다. Quant align Research에 따르면 전 세계 디지털 정신력 헬스케어 시장은 2027년에 200억 달러 규모까지 성장할 것으로 전망했다[38].

MOST 시스템[39]이 짧은 대화를 통해 사용자를 관련된 정신 건강 콘텐츠로 연결하는 것을 시작으로 정신 건강 AI인 Wysa[40], Tess[41]가 등장하였다. Wysa는 인지 행동 치료와 함께 행동 강화를 통해 우울증 치료를 돕는다. Tess는 정신 건강 문제로 어려움을 겪는 사용자들과 대화와 공감을 나누는 챗봇이다. 위 세 챗봇은 공통으로 일상 대화를 시작으로 그들에게 일어난 이벤트, 기분을 끌어내 원인을 파악하고 정신 건강을 향상하기 위한 간단한 치료와 의학 정보를 제공하는 챗봇이다. 이렇게 대화 수준의 대화형 인공지능을 통해 정신건강 분야에 접근한다면 정신 질환의 조기 발견, 치료 접근성 감소 등의 이점을 끌어낼 수 있다.

대표적인 선도 기업인 마인드스트롱(Mindstrong)[42]은 사용자의 스마트폰 사용 패턴을 수집해 인지 기능과 감정 상태를 측정하고, 이를 통해 원격 심리 상담

과 원격 진료를 제공한다. 스마트폰을 사용할 때 스크롤의 패턴, 타이핑 속도 등을 전통적인 인지기능 검사 방법과 연관 지어 사용자에게 개인화된 의료 서비스를 제공할 수 있다.

인지 행동 치료(CBT) 기반 상담 챗봇인 워봇(Woebot)[43]은 자연어처리, 심리학 지식, 글쓰기 능력과 유머 감각을 조합해 사용자에게 친근한 대화를 제공한다. 치료사보다는 친근한 말투로 유머를 섞어 대화하는 워봇은 ‘한 가지만 보고 모든 걸 판단하는 것은 아니지?’와 같은 질문을 통해 사용자가 자신의 부정적인 생각을 스스로 되돌아볼 수 있도록 돕는다.

최근에는 비가시적인 상대와의 간단한 수준의 채팅이 아닌, 챗봇에서 실제 인간과 같이 캐릭터가 부여되거나 사용자가 가상의 본인을 만들어 챗봇과 대화함으로써 사용자가 보다 더 솔직하게 자신의 얘기를 털어놓고, 더 나은 품질의 치료로 유도하는 공감형 챗봇들이 등장하였다.

사용자가 자신의 챗봇을 위한 아바타를 직접 디자인할 수 있는 레플리카(Replika)[44]는 구글과 OpenAI의 텍스트 생성 딥러닝 모델을 사용하여 단순히 지정된 단어에 대한 답변을 하는 것이 아니라 사용자와의 대화 속에서 얻어진 사용자 경험의 축적을 학습해 진화하는 공감형 AI 에이전트이다. 레플리카와 같이 사용자가 가상의 아바타를 만들어 심리 상담 챗봇과 대화를 수행하면 스스로 문제를 인식하거나 해결하기 힘든 정신 질환 환자들이 챗봇 치료사와 활발히 상호작용 함으로써 효과적인 정신건강 서비스를 이용할 수 있다.

4. 정신건강 분야를 위한 공감형 AI 에이전트 활용 방안

정신건강 분야를 위한 공감형 AI 에이전트기술은 매우 다양한 분야에 적용하여 활용이 가능하다. 그림 2에서는 10가지의 정신 장애 특성과 AI 적용 분야에 따라 제공할 수 있는 서비스를 구분하였다 [45].

음성, 영상, 사진 인식 데이터와 챗봇과 사용자가 대화를 통해 교류하며 학습·추론 한 정보, 라이프로그, 생체, 뇌파 데이터 등 복합적인 데이터를 활용하여 정신건강을 위한 공감형 AI 에이전트 기술 기반 챗봇 서비스를 제공할 수 있다. 예방, 진단 지원, 원격 진단, 소통 지원, 감성 교류, 디지털 치료제, 감성 코칭, 치매 치료, 감정 조절, 돌봄과 같이 일상생활과 의료 기관에서 활용할 수 있는 서비스를 낮은 접근성과 함께 제공함으로써 소외된 지역이나 계층은 사람들의

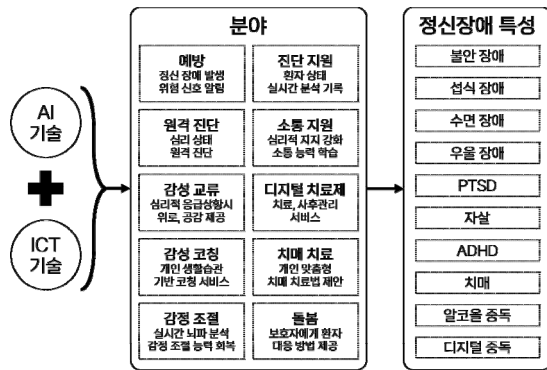


그림 2 정신장애 특성 및 AI 적용 분야

정신건강을 모니터링하고 적절한 치료를 보다 빨리할 수 있다. 또한 누구나 동일한 조건으로 서비스에 접근할 수 있기 때문에 기업이나 기관에서 직원들에게 맞춤형 건강관리 제공이 가능하다.

메타버스 기반 공감형 챗봇을 통한 정신 건강 관리 역시 기대를 받고 있다. 현재 스페인의 사이어스(Psious)는 VR/AR 기술로 정신 장애를 야기하는 상황과 자극에 노출하는 방식으로 심리 치료에 적용하고 있다[46]. 단순한 VR/AR을 넘어 메타버스로 확장한다면, 기기나 물리적 공간을 전환하지 않고도 메타버스 공간에서 사용자의 행동 패턴, 실제 사용자의 신체, 생활 데이터, 공감형 챗봇과의 상호 작용으로부터 얻어진 데이터 등을 복합적으로 활용해 메타버스 내에서 치료를 제공할 수 있다. 앞서 언급했던 대로 현대인들이 정신 질환 치료를 잘 받지 않는 이유에 거부감과 두려움이 큰 부분을 차지하고 있었다. 그러나 메타버스 공간 내에서 제2의 자신을 만들어 정신과라는 실체적 공간에 가지 않고도 일상 대화를 통해 편하게 증상에 관해서 얘기하며 자기 방어적 태도를 최소화할 수 있다는 장점이 있다. 따라서 메타버스 공간에서 공감형 챗봇을 적용한다면 사용자와의 솔직한 공감형 대화와 함께 여러 센서로부터 얻은 데이터의 효율적인 사용이 가능하다.

5. 결 론

본 원고에서는 정신건강 분야를 위한 공감형 AI 에이전트를 구축하기 위한 주요 기술과 국내외 기술 동향을 분석하여 정리했다. 정신건강 분야의 AI 에이전트는 기존 범용 AI 챗봇과 달리 텍스트, 이미지, 라이프로그 등 멀티모달을 기반으로 공감 기반의 대화가 필수적이며 전문적인 지식 기반의 대화를 제공할 수 있어야 한다. 또한 이를 기반으로 활용할 수 있는 방안에 대해 논의함으로써 정신건강 분야의 기존 챗봇

들의 한계점을 극복하고 정신건강 분야에 대한 접근성을 높이고 전문 인력을 지원할 수 있을 것으로 기대한다.

정신건강 문제가 사회적으로 매우 중요한 이슈로 논의되고 있기 때문에 AI 기술을 활용하여 이를 극복하기 위한 연구가 활발하게 진행될 필요가 있다. 여전히 민감한 개인정보 활용 문제, 신뢰 가능성 문제 등 해결해야 할 부분이 존재한다. 사회적으로 중요한 이슈인 정신건강 문제를 해결하기 위해서는 결국 기술뿐만 아니라 전문 지식과 정책이 필수적이다. 정부의 개입을 통한 정책과 정신건강 의학 분야와 AI 기술의 융합을 통해 지속적으로 연구 및 개발을 수행한다면 정신건강 분야에서 공감형 AI 에이전트의 적용 범위가 확대될 수 있다. 공감형 AI 에이전트의 적극적 활용으로 부족한 전문 인력을 지원하고 정신건강 분야에 대한 접근성을 높임으로써 사회적 문제에 있어 긍정적인 효과를 기대할 수 있다.

약어정리

- EEG (ElectroEncephaloGraphy)
- ADHD (Attention Deficit Hyperactivity Disorder)
- PTSD (Post-Traumatic Stress Disorder)
- CBT (Cognitive Behavioral Therapy)
- PHQ-9 (Patient Health Questionnaire-9)
- Seq2Seq (Sequence to Sequence)
- CNN (Convolution Neural Network)
- GAN (Generative Adversarial Networks)

참고문헌

- [1] HIRA 빅데이터 개방 포털, 입원/외래별 국민관심질병 통계-우울증(2021), 2021.
- [2] 보건복지부 보도자료, “고립·은둔 청년, 이제 국가가 돕겠습니다.”, https://www.mohw.go.kr/board.es?mid=a10503000000&bid=0027&list_no=1479278&act=view, 2023.
- [3] 국립정신건강센터 정신연구소, “정신건강 인식개선, “MIND, No. 4, 2023.
- [4] Kocon, J. et al., “ChatGPT: Jack of all trades, master of none.”, Information Fusion, No.101861, 2023.
- [5] Google, “An important next step on our AI journey”, <https://blog.google/technology/ai/bard-google-ai-search-updates/>, Feb 06, 2023.
- [6] Rosales, J. H., Rodriguez, L. F., Ramos, F., “A general theoretical framework for the design of artificial emotion

- systems in autonomous agents.”, Cognitive Systems Research, Vol.58, pp.324-341, 2019.
- [7] O’Shea, K., Nash, R., “An introduction to convolutional neural networks.”, arXiv preprint, arXiv:1511.08458, 2015
- [8] Yu, Y., Si, X., Hu, C., Zhang, J., “A review of recurrent neural networks: LSTM cells and network architectures.”, Neural computation, Vol.31, No.7, pp.1235-1270, 2019.
- [9] Vaswani, A., Shazeer, N., et. al., “Attention is all you need.”, Advances in neural information processing systems, Vol.30, 2017.
- [10] Devlin, J., Chang, M. W., Toutanova, K., “BERT: Pre-training of deep bidirectional transformers for language understanding.”, arXiv preprint, arXiv:1810.04805, 2018.
- [11] Zhang, W., et. al., “Transformer-based multimodal information fusion for facial expression analysis.”, In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp.2428-2437.
- [12] Wang, Y., et. al., “Multimodal token fusion for vision transformers.”, In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp.12186-12195.
- [13] Budzianowski, P., Vulić, I., “Hello, it’s GPT-2--how can i help you? towards the use of pretrained language models for task-oriented dialogue systems.”, arXiv preprint, arXiv:1907.05774, 2019.
- [14] Liu, X, Zheng, Y. et al., “GPT understands, too.”, AI Open, 2023.
- [15] Wang, Y. Yao, Q. et al., “Generalizing from a few examples: A survey on few-shot learning.” ACM computing surveys(csur), Vol.53, No.3, pp.1-34, 2020.
- [16] Huang, H. Yu, P. S. Wang, C., “An introduction to image synthesis with generative adversarial nets.”, arXiv preprint, arXiv:1803.04469, 2018.
- [17] AI Hub, “Wellness dialogue script datasets.”, Available at <https://aihub.or.kr/node/267#>, 2019.
- [18] Liu, B., Mazumder, S. et. al., “Lifelong and continual learning dialogue systems: Learning during conversation”, 35th AAAI Conference on Innovative Applications of Artificial Intelligence, Vol.35, No.17, pp.15058-15063, 2021.
- [19] 유소엽, 정옥란, “BERT 모델과 지식 그래프를 활용한 지능형 챗봇.”, 한국전자거래학회지, 제 24권, 제 3호, pp.87~ 98, 2019.
- [20] 박채림, 유소엽, 정옥란, “연속학습 기반의 대화형 AI agent.”, 한국컴퓨터정보학회논문지, 제 28권, 제1호, pp.27~ 38, 2023.
- [21] Ahn, Y. S., Jeong, O. R., “Time-Aware PolarisX:Auto-Growing Knowledge Graph.” Computers, Materials & Continua, Vol.67, No.3, 2021.
- [22] Dale, R., “The return of the chatbots.”, Natural Language Engineering, Vol.22, No.5, pp.811-817, 2016.
- [23] Kulkarni, P., Mahabaleshwarkar, A., Kulkarni, M., Sirsikar, N., Gadgil, K., “Conversational AI: An overview of methodologies, applications & future scope.”, in 2019 5th International Convergence On Computing, Communication, Control And Automation(ICCUBEA), IEEE, pp.1-7, 2019.
- [24] Weizenbaum, J., “ELIZA-a computer program for the study of natural language communication between man and machine.”, Communications of the ACM, Vol.9, NO.1, pp.36-45, 1966.
- [25] Batacharia, B., Levy, D., Catizone, R. et. al., “CONVERSE: a conversational companion.”, In Machine conversations, Springer US, pp.205-215, 1999.
- [26] Pamungkas, E. W., “Emotionally-aware chatbots: A survey.”, arXiv preprint, arXiv:1906.09774, 2019.
- [27] Hien, H.T., Cuong, P.-N. et. al., “Intelligent assistants in higher-education environments: the FIT-EBot, a chatbot for administrative and learning support”, In Proceedings of the Ninth International Symposium on Information and Communication Technology, ACM, pp. 69 - 76, 2018.
- [28] Sanderson, K., “GPT-4 is here: whar scientists think.”, Nature, Vol.615, No.7954, pp.773, 2023
- [29] Gemini team, “Gemini: A family of highly capabel multimodal models.”, Google, 2023.
- [30] Xu, Xing, et al., “Cross-modal attention with semantic consistence for image-text matching.”, IEEE transactions on neural networks and learning systems, Vol.31, No.12, pp.5412-5425, 2020.
- [31] Rothbaum, B. O. Meadows, E. A. et al., “Cognitive-behavioral therapy.”, 2000.
- [32] Shah, H., “ALICE: an ACE in Digitaland.”, tripleC: Communication, Capitalism & Critique, Open Access Journal for a Global Sustainable Information Society, Vol.4, No.2, pp.284-292, 2006.
- [33] Zhou, H., Huang, M., Liu, B. et. al., “Emotional Chatting Machine: Emotional conversation generation with internal and external memory.”, In proceedings of the AAAI Conference on Artificial Intelligence, Vol.32,

No.1, 2018.

- [34] Li, J., Sun, W. et. al., "Reinforcement learning based emotional editing constraint conversation generation.", arXiv preprint, arXiv:1904.08061, 2019.
- [35] Sun, X., Chen, X., Pei, Z., Ren, F., "Emotional human machine conversation generation based on SeqGAN.", In 2018 First Asian Conference on Affective Computing and Intelligent Interaction, IEEE, pp.1-6, 2018.
- [36] Ghandeharioun, A., et al., "Emma: An emotion-aware wellbeing chatbot.", In 2019 8th International Conference on Affective Computing and Intelligent Interaction(ACII). pp1-7, 2019.
- [37] Kroenke, K. Spitzer, R. Williams, J. B., "The PHQ-9: validity of a brief depression severity measure.", Journal of general internal medicine, Vol.16, No.9, pp.606-613, 2021.
- [38] Quantailgnresearch - QuantAlign Research, "Global Digital Mental Health Market Demand Outlook, COVID-19 Impact, Trend Analysis By Deployment (Mental Health Chatbots, Tele psychiatry, Mindfulness & Meditation Apps, Mobile / Platform Based CBT, VR/AR Based Therapy), By Component, By Software Services, By End Use and Industry Estimates 2021-2027", 2021.
- [39] D'alfonso, S., Santesteban-Echarri, O., Rice, S. et. al., "Artificial intelligence-assisted online social therapy for youth mental health." Frontiers in psychology, Vol.8, pp.796, 2017.
- [40] Inkster, B., Sarda, S., Subramanian, V., "An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: real-world data evaluation mixed-methods study", JMIR mHealth and uHealth., Vol.6, No.11, e12106, 2018.
- [41] Rauws, M., Quick, J., Spangler, N., "X2 AI Tess: working with AI technology partners", 2019.
- [42] Metz, R., "The smartphone app that can tell you're depressed before you know it yourself.", Technol Rev, Vol.15, No.1, 2018.

- [43] Beiley, M. R., "Mental health and wellness Chatbot", 2019.
- [44] Possati, L. M., "Psychoanalyzing artificial intelligence: The case of Replika.", AI&SOCIETY, Vol.38, NO.4, pp.1725-1738, 2023.
- [45] 송근혜, 김문구, 박안선, "정신건강을 위한 인공지능 활용과 유망 서비스", 전자통신동향분석, ETRI, 제 35 권, 제 6호, pp.12~23, 2020.
- [46] McMahon, E., "Virtual reality exposure therapy: bringing 'in vivo' into the office", Journal of Health Service Psychology, Vol.43, pp.46-49, 2017.

약 력



고 효 진

2023 가천대학교 AI·소프트웨어학부 졸업 (학사)
2023~현재 가천대학교 AI·소프트웨어학과 석사과정
관심분야: 멀티모달, 지식 그래프, 정보 추출
Email : 2rhgywls@gachon.ac.kr



유 소 업

2014 가천대학교 소프트웨어설계·경영학과 졸업 (학사)
2016 가천대학교 소프트웨어설계·경영학과 졸업 (석사)
2021 가천대학교 AI·소프트웨어학과 졸업 (박사)
2021~현재 가천대학교 AI·소프트웨어학부 연구교수
관심분야: 대화형 AI, 지식 그래프, 공감형 AI
Email : bbusso@gachon.ac.kr



정 옥 란

2005 이화여자대학교 컴퓨터공학과 졸업 (박사)
2006 서울대학교 컴퓨터공학부 (박사후연구원)
2007 Univ. of Illinois Urbana-Champaign (박사후 연구원)
2008 성균관대학교 정보통신학부 (연구교수)
2017 Univ. of California, Irviene (방문교수)
2009~현재 가천대학교 AI·소프트웨어학부 (교수)
관심분야: 빅데이터, 머신러닝, 대화형 AI, 공감형 AI
Email : orjeong@gachon.ac.kr