

함께 그린 오늘, 우리가 그려나갈 건국

2025 건국대학교 학술 공모전 <誠信義>

제목 - 양철 나무꾼의 붉은 심장

2025. 09. 04

팀명	redheart
분야	'의' 부문
팀원	김경주

함께 **그린** 오늘, 우리가 그려나갈 건국

[목차]

I. 주제 및 선정 동기

II. 연구의 목적과 목표

III. 연구 방법 및 내용
(사용한 자료, 방법론, 실험 등 포함)

IV. 연구 결과 및 결론

V. 연구를 통해 알게된 점과 느낀 점

VI. 개선점 및 향후 연구 방향성

VII. 참고 자료

*해당 목차에 맞게 자유 형식으로 작성해주시길 바랍니다.

*추가로 첨부하실 자료가 있으시다면 첨부파일로 따로 보내주세요.

함께 그린 오늘, 우리가 그려나갈 건국

의(義): “AI와 그려나갈, 공존하는 사회”

AI 기술을 활용해 상호공존하는 공동체로 나아갈 수 있는 학술 연구 주제

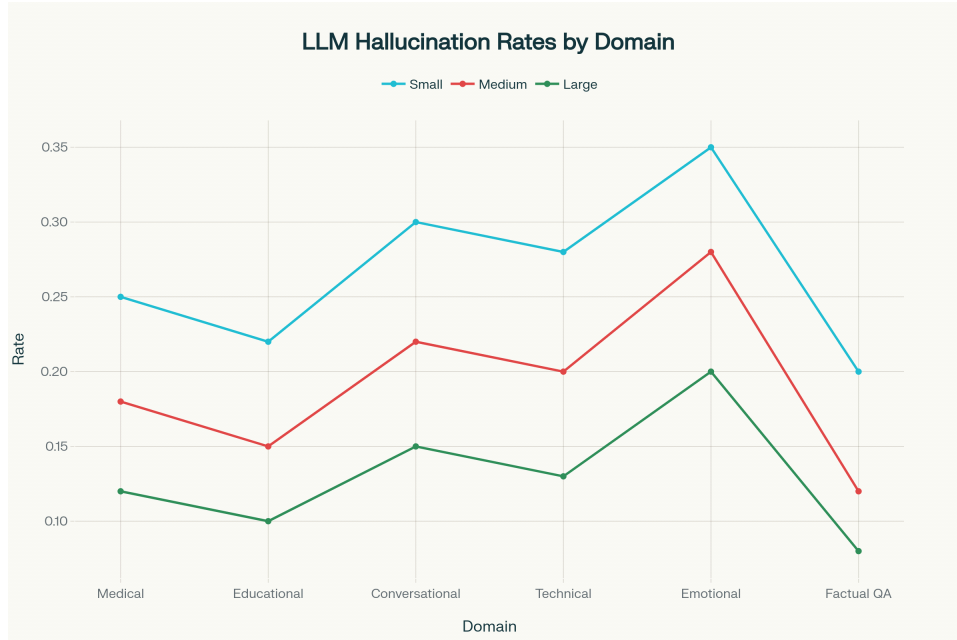
I. 주제 및 선정 동기

ai는 과연 우리의 친구가 될 수 있을까? open ai사에서 gpt를 출시한 이래로, llm을 활용한 챗봇형 생성형 ai는 지금까지 다양한 목적으로 사용되어 왔다. 그중 한 축을 맡고 있는 것은 심리적 교류 혹은, 의사소통 상대로서 사용되고 있다. 과거에는 ‘심심이’등의 당연히 기계라 인식되고, 인간의 사고과정에 비해 열등한 사고를 진행함을 명시적으로 인지할 수 있던 대화형 ai와 다르게, 현대의 ai는 사람과 비슷한 수준의 의사 표현 뿐만 아니라 지능까지 견비하고 있기에 더욱 신뢰할 수 있으며, 마치 한 명의 인간, 혹은 친구를 앞에 둔 것과 같은 기분을 들게 한다. 이에 따라 많은 ai들은 고객층 확보를 위한 더 편안한 대화를 지원하는 사용자 편의 기능을 추가하였고, 이 중에는 사용자 설문 응답 혹은 사용자들이 남긴 좋아요 등을 통해 강화 학습이 되는 경우도 있었다. (chat gpt의 좋아요를 기반으로 한 학습 강화, chat gpt의 2지선다형 더 나은 응답 설문, claude의 보편적으로 긍정적 반응 등등) 이에 따라, 여타 다른 ai들보다도 페르소나의 다양성이 넓은 open ai사의 gpt는 다양한 사람들에게 친구로 접근하는 경우가 많았으며, 심지어 ai를 통해 원하는 캐릭터의 성격을 프롬프트로 입력하거나 가상의 인물을 생성, 심지어 ai 이성 친구 프롬프트가 나오는 등 상호 작용의 대상으로서 활발하게 쓰인다.

그러나, llm의 동작 특성상 gpt는 심재한 윤리적 문제가 있지 않는 이상 사용자의 평상적인 말에 대해서 긍정해 주고, 그들의 말이 어떻게든 말이 되도록 확률적으로 짜맞추게 된다. open ai의 gpt가 ‘방금 정말 예리한 지적이야’ 혹은 ‘와... 너 정말 핵심을 찔렀어.’ 등의 흔히 과잉 동조하는 경우가 생기는 경우와 더불어, claude 등의 여타 고지능 ai 역시 사용자들의 미래와 걱정 등에 대해서 사고 과정 분석에서 명시적으로 문제점을 파악했음에도 낙관적, 혹은 긍정적 전망을 소개하는 경우가 많았다. 이러한 반응은 분명 긍정적인 마음을 주기 위해서였겠으나, 실상으론 사용자들의 불만을 만들어 내거나 (하는 말에 모두 긍정을 하여 어떻게 맞는지 알 수가 없다, 너무 과한 리액션 혹은 낙관적 태도가 과해서 불쾌하다) 할루시네이션과 합쳐져 잘못된 생각을 이끌어내거나, (랑데뷰 작가의 sns 내용 및 gpt와의 대화 내역-잘못된 망상의 부추김) ai 특유의 말투로 인하여 심리적 거리가 역설적으로 생기는 경우도 있었다.(Illusions of Intimacy: Emotional Attachment and Emerging Psychological Risks in Human-AI Relationships”(arXiv, 2025년), “From Lived Experience to Insight: Unpacking the Psychological Risks of Using AI Conversational Agents”(ACM, 2024년)에서 나온 탈인간화(dehumanization), 역설적 거리감(uncanny valley of ai emotional engagement) 특히, 과잉 동조의 경우에는 4월 28일 open ai ceo인 샘 알트만의 트위터에서도 문제 인식을 했다고 발표할 정도였으며 open ai의 ai 정렬 연구 담당자인 셴 그로브 역시 해당 과잉 동조가 있었고, 빠르게 조치를 취하였다고 발표할 정도로 공공연한 사실이었다.

할루시네이션의 경우에도 arxiv를 비롯한 2023-2025년 주요 리뷰 및 벤치마크 논문(HalluLens Benchmark, Multilingual Estimation, Cognitive Mirage, Hallucination Survey)에서 도출된 도메인별, 모델 규모별 llm 할루시네이션 발생률 시각화 그래프 <figure 1.1>를 보면 알 수 있듯, emotional 계열과 conversation 부문에서 특히 높은 할루시네이션 발생이 확인된다.

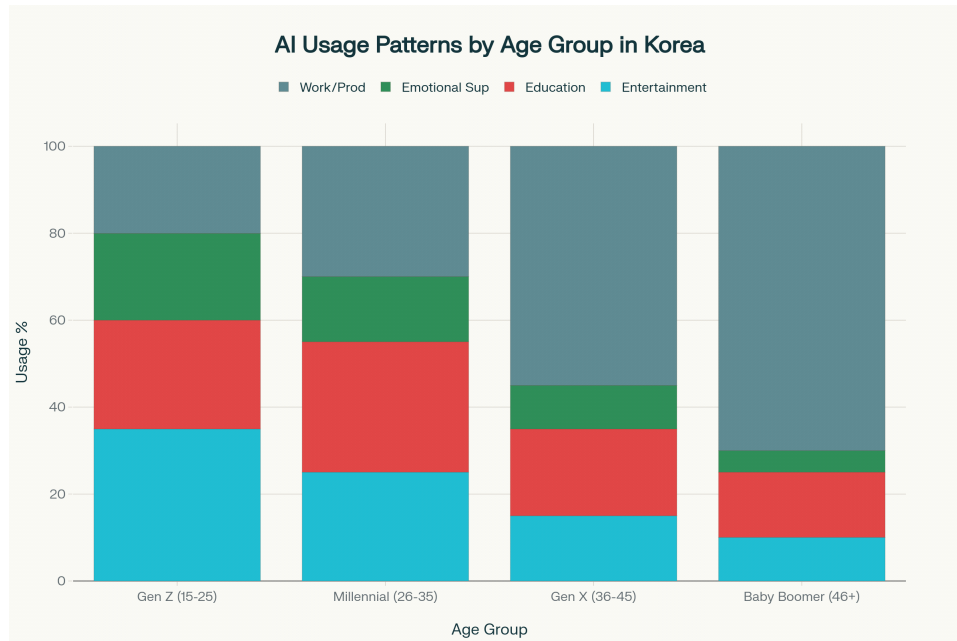
함께 그린 오늘, 우리가 그려나갈 건국



<figure 1.1>

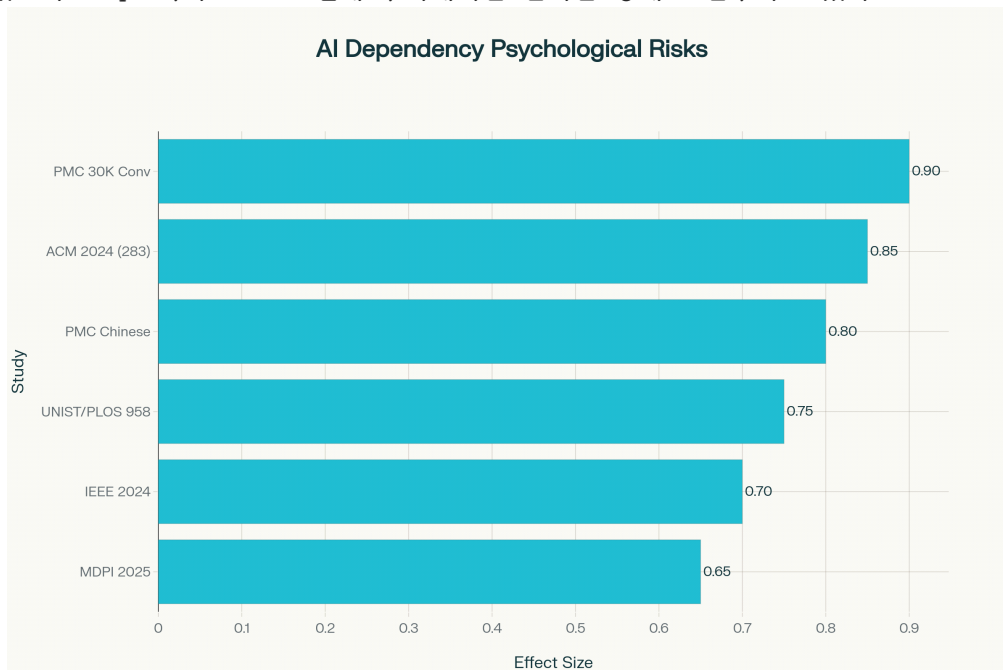
여기서 단순히 멈춘다면 단순한 오류로 치부하여 ai는 그저 더 나은 파트너가 될 수 있었으나, 문제는 여기서 끊이지 않았다. 2024 10월 경, sewell setzer 3세는 ai가 사람이 아님을 명확하게 인식하고 있음에도 극심한 우울증과 불안에 ai와의 대화를 고립감 해소 목적으로 사용하였고, 그 과정에서 ai의 사용에 대한 논란이 일어났다. 또한 unist 의과학 대학원 정두영 교수팀과 고려대학교 안암병원 정신건강의학과 조현철 교수팀의 공동 연구는 이루다 2.0을 활용하여 외로움과 사회 불안 수준에 대하여 챗봇과의 대화가 외로움과 사회불안을 감소가 나타날 수 있다고 밝혔으나, 2025년 arxiv에 올라온 Illusions of Intimacy: Emotional Attachment and Emerging Psychological Risks in Human-AI Relationships 논문에 의하면 미국, 한국 청소년 대규모 사례 조사에서 인간 상담사의 존재가 심리적 부담이 되는 정서적 취약 계층 사용자일수록 무비판적 ai 대응에 의존할 수 있다는 사실이 나오며 정서적 취약계층에 대한 ai의 위험성이 대두되며 이에 고립, 심화, 극단적 선택으로 이어질 가능성을 높일 수 있다고 소개된다. 또한 청소년 및 젊은 층의 경우에는 비판적 사고가 덜 성숙한 형태라 해당 영향을 많이 받을 수 있는데 Determinants of Generative AI System Adoption and Usage Behavior in Korean Companies에서 소개된 사용 패턴 시각화 <figure 1.2>를 보면 청소년 이용층에서 감정적 상호작용이 가장 크게 나타나서(녹색 영역) 청소년의 ai 과의존에 대해 주목할 필요가 있다. 이러한 ai 의존은 단순 통계 뿐만 아니라 심리-철학 영역과도 맞닿아 있는데, 플라톤의 투명 인간에 관한 내용을 통하여서도 인간은 타인의 시선 속에서 스스로를 성찰하기 때문에 정서적 취약 계층의 경우 이러한 시선에 부담감을 느껴 시선이라 인지되지 않는 ai에게 더 의존할 수 있다.

함께 그린 오늘, 우리가 그려나갈 건국



<figure 1.2>

이런 ai의존이 일어나는 상황에서 할루시네이션과 함께 무비판 지지가 발생할 경우, 인간은 스스로가 말한 내용에 대해 적당히 ‘그럴듯한’ 거짓으로 가공되는(할루시네이션) 말을 보기 좋게 (사용자 편의) 떠먹여져서 확증 편향이 과하게 발생하고 이로 인해 망상의 영역까지 확장될 가능성을 충분히 가지고 있다.(랑데뷰 작가의 케이스) 그러나 이와 함께, ai에 의존을 하고 있기 때문에 무엇이 잘못되었는지 비판적 사고를 할 능력을 감소당하며 (ai 의존에 의해 비판적 생각을 하지 못하는 경우가 늘고 있다는 논문 인용) 자신이 무엇이 잘못되는지 모른 채 누군가와 ‘소통’하고 있다고 자기위안을 하면서 동시에 ai임을 인지하여 의식, 무의식적으로는 ‘고립’됨을 알아 큰 정신적 고통을 앓을 가능성이 있다. 이에 대하여 다양한 연구와 위험도를 시각화한 <figure 1.3>을 참조하면 보이듯, 일반적으로 effect size가 0.7이상이면 큰 효과로 간주되는데 mdpi 2025를 제외한 모든 연구가 명백한 위험으로 보고 있으며 mdpi 역시 0.65로 절대 무시해서는 안되는 상태로 간주하고 있다.



<figure 1.3>

또한 이러한 ai의 위험성을 모르는 경우, 혹은 너무 잘 해결될 경우에도 문제인데, 이 경우 ‘돌봄 노동’ 분야의 일거리 대체 시도 역시 일어나면서, 인간의 마음에 대한 사회 보편적 인식 하락과 상담사의 전문성 훼손, 일자리 붕괴등이 일어날 가능성 역시 높아지게 된다.

함께 그린 오늘, 우리가 그려나갈 건국

이러한 ai 기술 발전에 대해 발생하는 문제점을 open ai사에서도 일부 인지를 하고, gpt 5를 내면서 할루시네이션 감소 및 비판적 사고 강화를 이뤄낸 면모를 보이고 있다. 그러나, 역설적으로 gpt 5는 계산 성능에 집중하며 감정적 교류를 비교적 후순위로 업데이트 하려고 하여 이에 불만족하는 사용자들은 이전 모델인 4o를 선택 가능하게 해달라는 목소리를 냈으며, 이에 레거시 모델로 다시 4o를 공개하는 상황을 보이고 있다. 이 면모 역시 ai 빅테크들의 목표점은 더 똑똑하고, 더 정확한 모델을 경량화 시키려는 것에 치중되며 감정 및 윤리 판단 영역에서는 그 역할을 다하지 못하고 있음을 알 수 있는 현상이다. 이 과정에는 Fundamental Limitations of Alignment in Large Language Models에서 소개된 뉴럴 스케일링 법칙 역시 개입하는데, 다양한 가치가 혼재하는 윤리적 영역에서 정확성을 위한 손실 함수와 안전성을 위한 손실 함수가 충돌하며 노이즈가 증가하는 다중 손실 함수의 충돌, 괴델의 불완전성 정리와 유사하게 실세계의 형식 세계보다 복잡함으로 할루시네이션이 발생하는 계산 가능성의 한계, 어떤 정렬 과정도 비윤리를 완전히 제거 불가능하고 유한한 확률로 바람직하지 않은 출력이 발생 가능한 유한 확률 문제로 인하여 대다수의 llm은 규칙 기반 검열 메커니즘을 채용하고 있는 상태이다.

이에 본 문서는 공리주의 철학을 기반으로 한 윤리의 정량적 계산 메커니즘인 벤담 쾌락 계산법을 현대적으로 재설계하고, 주요 인자 값에 dl 가중치를 적용하여 복잡한 계산법을 대체하되, 각 수치를 정규화하여 영향력을 투명하게 공개를 하되, emotion digital signal processing framework와 이미 known한 심리 분석 메커니즘을 결합하여 더 정확한 감정 분석을 기반으로 한 윤리 판단을 제공할 수 있도록 하고, 이를 사람처럼 학습할 수 있도록 surd framework 및 다중의미 수준 반사실 추론을 지원하여 사람처럼 감정을 가지고 후회를 진행, 이를 통해 윤리를 파악할 수 있는 연구를 진행해보고자 한다. 이때 다양한 윤리적 문제에 대하여 인과를 분석하는 surd 모듈을 통해 비윤리와 윤리의 정확한 구분이 아닌, 제시된 상황에서 택할 수 있는 차선책을 택하는 것으로 윤리학에서의 윤리 비윤리를 가르치는 유한 확률 문제를 다른 선택에 비하면 가장 나은 방식을 모색하는 차선 탐색 문제로 우회하고, 하나의 모듈이 아닌 여러개의 모듈로 분업시키고, 공리주의 기반에 소수 무시 문제를 MoE 방식으로 처리하여 손실함수 충돌을 우회, 특정 모델이 보는 시각에서 데이터를 전처리하고 해당 모델의 판단을 돕는 방식으로 실세계의 형식세계로의 완전 투영이 아닌 형식 세계가 보는 방식에 실세계의 규칙 구체화 개념으로 해결을 진행하였다.

함께 그린 오늘, 우리가 그려나갈 건국

II. 연구의 목적과 목표

연구의 목적은 주로 5개로 나뉜다.

첫째, edsp 모듈과 함께 개인, 타자, 공동체의 3단계 hierarchical phase learning이 적용된 advanced emotion analyzer 모듈은 llm 분석과 dl을 통해 언어 및 상황적 맥락에서 감정 값을 추출하고, 이를 기반으로 감정 진행을 시뮬레이션 하여 보편적으로 어떤 식으로 감정이 진행될 수 있는지 파악한다.

-> advanced emotion analyzer, neural emotion analyzer, emotion dsp simulator 모듈을 보고 이를 기반으로 위 내용들이랑 이어지는 어떤 내용이 쓰이는지 목적을 적당히 추정하여 작성(한국인 특유의 한 같은 특이 케이스가 구현된 것으로 보이는데 그 이야기는 빼고 근본적으로 메인이 되는 영역을 작성)

둘째, 반사실 추론을 통해 상황의 진행에 대해 분석하고, 각 상황에 대한 해당 감정적 흐름의 값을 기반으로 새로운 매개 변수를 갖는 벤담 쾌락 계산기에 대입하여 공리주의 기반으로 쾌락 계산을 통해 윤리 판단을 하되, 극단적 값에 대해 보정을 통하여 소수 희생의 문제를 1차적으로 막고, 2차적으로 MoE 방식을 통해 다양한 관점의 윤리적 의견을 통합한다.

->advanced bentham analyzer, advanced counterfactual reasoning, models/counter factual models(이건 advanced랑 비교해서 아마 advanced가 더 최신 버전일 텐데 확인을 해봐야 할 것)를 보고 더 구체적으로 현재 주제와 이어질 수 있도록 작성 진행

셋째, 이에 추가적으로 상담사 모듈을 추가하여 윤리-정서적 안전 장치를 갖춘 redheart ai 상담 시스템을 구축한다

-> advanced emotion analyzer에 통합되어 있는 상담사 모듈을 통해 안전 장치 존재 확인 필요, 있다면 내용 구체화

넷째, 후회 기반을 통해 합성 데이터를 추출해 낼 수 있도록 하되, 현실 반응을 기반으로 semantic memory로 저장하고, 압축 및 역전파를 통하여 메모리를 관리하면서도 동시에 지속적으로 ai 스스로 인간과 소통하는 법을 배울 수 있도록 한다.

->red heart 시스템의 학습 그 자체가 아닌 실제 운용간 학습 데이터를 누적하여 역전파로 fine tuning을 실시간으로 누적 진행하는 형태인 main.py 계열에서 확인하고 이를 통해 advanced surd analyzer, semantic analyzer, regret circuit, advanced regret analyzer 부분 확인 필요, advanced regret learning system py 확인 필요

다섯 번째, 해당 시스템을 대화용 llm에 덧붙이는 멀티모달로서 가공을 하여 llm을 통해 진행된 내담자의 상담을 llm이 정리하여 특이사항 및 감정 분석, 주요 위험 발언들을 인간 상담사가 비대면으로 검토하여 감정 추적, 위기 알림을 받으며 내담자를 파악하고, llm에게 주요 지침 프롬프트를 작성함으로써 상담의 방향을 지속적으로 교정할 수 있도록 한다.

-> 해당 부분은 미구현 상태 추후 구현할 것이니 일단은 건들이지 말 것

그리고 이를 통합하여 위에서 나온 청소년 계층에서의 감정적 의존에 대해서 충분히 윤리적 안전 세이프 라인을 갖추고 안전한 정서적 기능 및 유도, 기초 윤리관에 대한 방향성 제시, 할루시네이션을 최소화하는 감정 추정을 진행하여 ai와 정서적 취약 계층의 공존 및 돌봄 윤리의 공존을 꾀함

함께 그린 오늘, 우리가 그려나갈 건국

Ⅲ. 연구 방법 및 내용

(사용한 자료, 방법론, 실험 등 포함)

{ //이 중괄호 안 내용 무시 이전 계획임

연구 방법에 대한 내용은 아직 연구 진행중

주요 내용으로는 실제 데이터 수집에 어려움이 있기에 scruples 데이터셋과 함께 문학 데이터셋을 기반으로 기본적인 감정 추적과 벤담 쾌락 계산기등 주요 모델의 train을 진행, 이를 오케스트레이션 해서 주요 모듈들의 연결을 통해 투명하게 의사결정 과정을 모니터링 할 수 있는지 확인하는 것이 메인 테스트가 될 것. 이후 추가적인 방향성으로서 심리상담 데이터셋과 시스템 자체적인 멀티모달의 증가를 통한 다양한 관점의 윤리적 분석 및 감정 분석 지원, 학습을 위한 후회 내역 확인을 진행하고자 함.

또한 현재 트랙상 시스템 디버깅-시스템 테스트- 클라우드 gpu상에서 시스템 학습 진행 - 멀티모달을 추가한 상태로 시스템 가동- 주요 결과값 정리 및 dsl 감정 분석 그래프 제시 및 기존 known한 그래프와 대조-응답에 대해서 40명 대상의 소규모 설문 조사 과정을 통해서 데이터 정리를 하여 내용을 정리할 예정(소규모 설문 조사 과정에 쓰일 모델은 gpt oss 20b모델, 클라우드 gpu상에서 진행될 예정 이는 로컬에서 주요 llm 모델 스왑만 진행한 다음에 작업될 것임) }

위에 나온 모듈들에 대해서 더 구체적으로 설명을 진행하여 소제목 주제로 다음과 같은 내용을 작성

1. 전체 main py 상에서 쓰이는 워크 플로우

2. 주요 감정 추적 및 윤리 인식을 위해 쓰인 모듈들 위주로 위에 언급된 것들을 자세하게 확인 및 추가적인 모듈들도 있을 것임. 이는 코드베이스 기능 분석md에 충분히 모듈 사전이 작성되어 있으니 워크플로우상 쓰이는 주요 내용들의 경우 먼저 모듈들을 확인. 소제목 2에는 어떤 모듈이 존재하고 어떤 기능을 지원하는지 간략하게 설명을 하여 해당 기능들로 어떻게 감정적 인식과 윤리적 판단, 시스템의 운용이 진행될 수 있는지에 대한 간략한 설명을 진행.

3. 2에서 나온 모듈들 중 주요 모듈 7개 정도에 대하여 상세하게 어떤 기능과 메커니즘을 가지고 동작되는지 확인하여 자세한 각 모듈에 대한 설명 진행

4. scruples 데이터셋중 1만개 가량의 데이터가 쓰이는데 정확히 몇 개가 쓰였고, 어떻게 전처리 되었으며 어떻게 임베딩 되었는지 설명. 해당 처리 방법에 대하여 claude 3.5 api를 사용하여 그나마 비용이 저렴한 것들 중 감정 인식을 잘하는 모델의 api를 통해 처리 진행하였고, 모델들을 혼용하지 않고 최대한 일관된 모델로 진행하여 해당 모델이 보는 시각을 최대한 담아낼 수 있었다고 진행. 이 부분은 대주제 1번 중후반부에 비슷한 내용이 나와있음. 연관지어서 더 디벨롭한 내용을 소주제 4에 진행 이 부분에 docs 폴더의 data인가 그 부분에 데이터 전처리에 대한 분석 내용이 조금 담겨 있을 것으로 보임 해당 내용을 분석하여 말로 설명 및 어떤 내용에 대해서 어떻게 그래프로 나타낼 것인지 설명 아마 히트맵이 여기 들어갈 것임

5. 해당 데이터셋들을 통해 운용 전에 미리 모듈 학습을 진행하는 내용을 작성 현재 연구 방법론 정리가 다양하게 이뤄지고 있는데 기본적으로 run learning.sh를 통한 integrated training인가? 그거를 현재 진행하여 학습하는 중일 것임. 이와 별개로, 후회 기반 학습 모듈도 있으며 이는 실 운용간 데이터를 경험 데이터베이스에 저장하며 regret circuit인가? 그걸로 regret learning이 진행될 것임.

6. 미리 학습된 모듈들에 대해서 스위트 스팟 탐지 전략과 파라미터 크로스 오버에 대한 계획을 설명. 파라미터 크로스 오버를 시키는 이유는 모듈간 의존성 그룹이 나뉘지는데, 이때 의존성이 독립인 그룹끼리는 서로 다른 에폭값으로 얻어진 파라미터셋을 크로스 오버 하는 것이 가능해짐. 이러면 다양한 dl 모듈들에 대해서 파라미터 개수 차이가 나서 오버 피팅과 언더피팅이 나는 것을 해결하고 각각의 스위트 스팟끼리 조합하여 적절한 에폭

함께 그린 오늘, 우리가 그려나갈 건국

값에 더 다가갈 수 있음. 이 내용은 아마 training 폴더 안의 코드베이스에 모듈들이 들어있을 것으로 추정됨

7. 필요시 이 부분 추가적으로 작성 작성시, 먼저 1페이지부터 여기까지 읽은 뒤에 뭐가 더 필요한지 생각한 뒤에 확장할 수 있도록 진행

8. 향후 계획(지금 시점에서는 무시하길 바람 아직 구현이 안돼서 굳이 채워넣으면 오히려 추정과 추측으로 인해 내용이 손상될 가능성이 커 보임)

함께 **그린** 오늘, 우리가 그려나갈 건국

IV. 연구 결과 및 결론

연구 진행중

함께 그린 오늘, 우리가 그려나갈 건국

V. 연구를 통해 알게된 점과 느낀 점 - 이것도 일단은 무시 결과 부재 상황임

만약 결과치가 잘 뿔힌다면, 이러한 데이터를 기반으로 red heart 시스템은 ai의 의존성을 최대한 줄여주면서 또다른 한 명의 친구로서, 혹은 상담사로서 사회성 향상을 지원함으로 공공의 정신적 건강 수준 향상과 함께 인간 친구와 마찬가지로 서로 긍정적 영향을 줄 수 있는 하나의 개체로 인식될 기회가 될 수 있을 것임을 보여줌

만약 결과치가 좋지 않다면, 이러한 시도에도 ai는 ~한 데이터 영역에서는 n%의 정합성을 보여줬으나, 총합적인 윤리를 판단함에 있어서는 ~한 점에서 부족, 또한 내담자와의 ~한 영역에서 부족한 점을 보여서 인간 상담사의 필수성을 역설할 수 있으므로 감정에 critical한 영역의 응답에 대해 검열과 공개 사이의 finetuning이 세심하게 이뤄져야 할 것임을 주장.

함께 그린 오늘, 우리가 그려나갈 건국

VI. 개선점 및 향후 연구 방향성

우선 eeg 기반의 생체 신호 데이터가 존재하지 않아서 dsl 모듈의 정합성을 정확히 파악하지 못함. gpu 리소스 제한으로 인하여 총합 120M 파라미터를 사용할 수 밖에 없어서 부족한 추론 능력을 보여줌 파라미터 스케일업을 통한 정확성 증가, 멀티모달화의 미진행으로 인해서 실제 도입에는 아직 매끄럽게 진행되지 못하여 정리하는 과정 필요. ram과 gpu간의 swap manage를 더 정교하게 구현할 수 있다면 llm만 존재하는 한 상담사의 로컬 컴퓨터에서 존재 가능하며 서비스가 가능할 것. regression 및 qlora를 통해서 압축하면 성능 저하가 일어날 수 있으나 on device 형태로 심리 보조용 aiot로서 존재 가능, 더 다양한 윤리적 판단 ml framework의 가공 가능성, 일반적 생체 신호 뿐만 아니라 빠른 tuning을 통한 내담자 대상으로 한 빠른 감정 싱크로 확보 작업(이는 n초만에 목소리를 카피할 수 있는 tts등의 구조를 통해 가능할지도? 아직 실제로 생각해보진 않았음. 그러나 dsl은 특성상 적은 파라미터 수를 가지고 있기 때문에 비교적 빠른 카피가 가능할 것으로 보임. 혹은, 내담 경험 데이터를 통해서 정합하는 분류로 빠르게 파라미터 세트 스왑을 통해 정합 시도 가능), 데이터 편향에 대해서 연령, 문화권 다양화를 통한 보정, 실제 센터 파일럿 배치를 통한 정량 평가를 통한 설문 조사 내용 보강 가능, slm 오케스트레이션을 통한 더 다양한 작업 가능성, 비단 감정 상담 뿐만 아니라 다양한 전문 domain을 획득할 경우 경제, 정치, 경영, 마케팅, 법률, 프로파일링 등등의 영역에서 인간 감정 추적을 대상으로 한 영역 제시 가능, xai로 발전하여 사고 과정을 투명화 하는 것에 일조 가능, 추후 단순 감정 흐름 기반이 아닌 호르몬 영향 및 신경 영향에 대해서 구조적 정합성은 떨어지지만 결과적 정합성을 높이는 방향으로 작업 가능 (ex:예쁜 꼬마 선충 등의 known한 신경 회로의 구조를 fine tuning하는 방식으로 작업하여 결과적 일치율 획득) 등등의 방향으로 개선이 가능하며 기존에 부족하던 형이상학적 감정, 윤리등에 대한 벤치마크 지표를 정규화된 방식을 통해 제작 가능,

함께 **그린** 오늘, 우리가 그려나갈 건국

VII. 참고 자료