Smilegate Membership AI 2기 신청서

1. 팀 소개

팀 이름	다크 레이디 (dark lady)	**인원 수**	3
팀 소개	**Q. 팀이 어떻게 결성되었나요?**		
	"우리의 다크레이디 (dark lady)가 다음 주면 여기를 떠난대" 다크레이디는 동료 과학자가 DNA 분자의 이중나선 구조의 X선 사진을 처음으로 발견한 로잘린드 프랭클린을 칭하던 별명이다. 다크레이디는 머리와 피부빛이 검은 여인, 사회통념적으로 예쁘지 않은 여인을 말하는 용어였다. 우리는 젠더와 인종 차별적인 용어 '다크 레이디'를 팀명으로 사용하여 앞으로 제 2의 다크레이디, 또는 제2의 다크 맨이 없는 차별없는 기술 사회를 만들기 위해 나아가려고 한다. 우리는 서강대학교 Art&Technology 전공 3명으로 이루어져 있으며 인문학에 기반한 스토리텔링과 상상력으로 과학기술과 예술을 융합하고 이를 통해 사회적 목소리를 낼 수 있는 시도를 하고 있다. <Natural Language Processing> 수업을 함께 수강하며 인공지능 분야 중에서도 자연어 처리 스터디를 진행하며 만났다. 그러던 중 현재 AI가 인종적, 젠더적 편견을 재생산하고 있는 문제와 버추얼 인플루언서가 소비되는 방식에 대한 공통된 고민을 나누었다. AI 엔지니어링을 공부하는 여성들로서 AI와 버추얼 인플루언서 시장에 현존하는 문제를 해결하고, 새로운 흐름을 만들어내기 위해 팀을 결성했다. 구체적 계획을 위해 인공지능 챗봇을 조사하던 중, 현재 나와있는 챗봇들에 문제점이 많다는 것을 알게 되었다. 특히, 불특정 다수에게 서비스를 제공하는 챗봇인 '이루다'가 최근 차별적, 혐오적 메세지를 표현하여 이슈가 되었다. 국내뿐만 아니라 국외의 인공지능 채팅봇들 사이에서도 잘못된 성적/인종차별적 언행을 하여 서비스가 중단된 사건이 발생하였다. 이는 인공지능이 편향된 데이터를 바탕으로 학습하기 때문이라고 생각했으며, 성과 인종, 지역을 포함한 데이터의 편향성으로부터 인공지능 챗봇의 결점이 생긴다는 것을 알게되었다. 팀 다크 레이디는 시중에 나와있는 AI 챗봇에 대한 불만과 갈증을 해소하기 위해 성,인종,지역에 있어 평등한 가치관을 가진 버추얼 AI 작가를 만들어보려고 한다. 최근 다양한 가상 인플루언서가 등장하고 있지만,		

대부분 여성과 엔터테인먼트 쪽으로 치우쳐있으며 마케팅의 요소로 사용되는 경우가 많다. 많은 사람들의 눈길을 끌어야 하는 광고의 특성에 맞추어 만들어진 대부분의 가상 인플루언서들은 사회적으로 아름답다고 여겨지는 편향된 미의 기준에 맞춰 제작되어 있어 일반 사람들의 공감대를 얻기 어려울 뿐더러 '불쾌한 골짜기'라는 수식어를 얻고 있다.

인플루언서를 자연스럽게 보이도록 하는 뛰어난 기술들에 비해 인플루언서 인격체에 대한 브랜딩이나 고민은 여성 서사가 중요시 되고 있는 시대적 흐름이나 변화에 늦어지고 있다고 생각한다. 샘다수는 가상 인플루언서에 성,인종,지역에 있어 평등한 가치관을 부여하여 잘나가는 브랜드의 가방을 메고, 유행하는 춤을 추는 보여주기식의 AI 인플루언서가 아니라 철학을 가지고 인간과 교감할 수 있는 AI 인플루언서를 만들어보려한다. 인플루언서의 의미를 넘어선 사회적으로도 선한 영향력이 되는 콘텐츠로 성장하길 기대한다.

<팀원 소개>

안지인은 기획과 기술 지원을 맡고 있다. 공연이나 프로젝트 기획에 관심이 있으며, 아이디어를 정리하고 계획을 수립하는 일에서 더 나아가 자신의 생각을 직접 표현하고 시각화할 수 있도록 프로그래밍을 공부하고 있다. 최근 AI와 머신러닝 분야에 흥미를 느껴 실제로 사용할 수 있는 서비스를 구현하는 프로젝트에 관심이 많다. 적극적으로 개발에 참여할 수 있도록 AI 분야에 대해 몸으로 직접 배울 수 있는 이 기회를 놓치고 싶지 않다.

- *(2020) 아트&테크놀로지 학과 전시회 'ATC2020' 전시 기획 팀원*
- *(2022) 중앙 밴드 동아리 '킨젝스' 기획*

장예원은 기획과 기술 지원(코딩 등), 디자인을 맡고 있다. 평소 과학과 철학, 과학과 인문학 사이의 연관성에 관심을 가지고 관련 공부와 작업을 하고 있다. 또한, 고립되고 개인적인 작품보다는 작업을 통해 관람객과 소통하길 원하며 보이지 않는 세계, 소외된 세계를 다양한 매체를 통해 표현하려는 시도를 하고 있다.

- (2019) 철학과 과학의 연관성을 담은 에세이 <하이젠베르크, 철학적으로 과학을 생각하다> 작성,

-제12회 노벨과학에세이 물리부문 동상

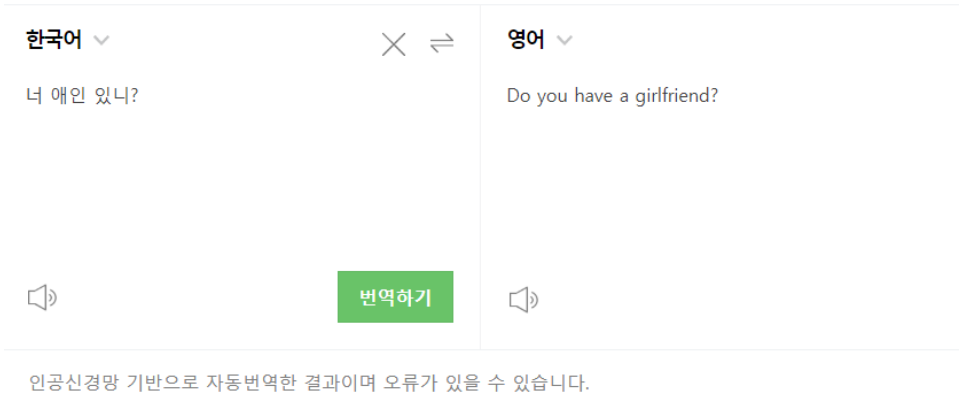
- (2021) 양자역학과 다중자아를 주제로 한 미디어아트 퍼포먼스 <겹과 결>, 기획 및 제작(Code art)

-2021 양자나노과학연구단 SPIN ART 공모전
양자나노과학연구단장상/유튜브 스트리밍상
-IBS와 서울예술대학교에서 주최한 <Beyond the Lens : Nano Bio Nature> 전시참여
- (2021) 포스트 코로나 시대의 행복을 주제로 한 3D 애니메이션 <BLOCK>, 3D Modeling & Animation
-제1회 중앙 미디어아트 공모전 우수상 (한국미디어아트협회 협회장 상)
-코엑스 미디어타워/파르나스 미디어타워에 동시 송출
- (2022) 대학생을 위한 메타버스 회의 플랫폼 <XTOWN> 디자인, 3D Modeling & UI/UX

이육샛별은 기획과 기술 지원을 맡고 있다. 숫자로 표현된 데이터가 사람들의 삶, 이야기와 연결되는 지점을 포착해 사회문화적 현상들을 데이터를 통해 표현하는 작업을 진행한다. 또한 평균적 인간에 맞춰진 데이터 분석과 AI 서비스가 아니라 데이터 분석을 통해 우리 눈에 보이지 않는 소외된 사람들의 이야기를 전하고, 다양한 존재의 사람들에게 맞춰진 AI 서비스를 개발하고 싶다.
-2020) 빅데이터 분석 Team Cayley 빅데이터 분석 프로젝트 <코로나와 우리의 기억> 사회팀
- (2020) 행정부에서 진행한 코로나 끝장개발대회 커뮤니티상
- (2020) 데이콘 법률안 NLP 분석 대회, N번방과 법률안 프로젝트
- 데이콘 금상 수상
- (2021) URL 관리 서비스, 개인사업자 세금 관리 웹 서비스 개발
- (2021) 멋쟁이 사자처럼 아이디어톤, 해커톤 전국 1위
- (2021) 아트&테크놀로지 학과 전시회 ATC 버추얼팀 리드, 메타버스 컨퍼런스 프로덕트 디자인

Q. AI 분야에 관심 갖게 된 계기는 무엇인가요?

AI 서비스는 이전에 존재하던 데이터를 기반으로 발전하기 때문에, 평소 우리가 미처 인식하지 못했던 사회의 문제가 드러난다는 특징이 있다고 생각한다.
이러한 문제에 대해 고민하기 시작한 계기는 한 포털사이트 번역기에서 우연히 한 문장을 돌려본 것이었다.



애인, 연인 등 어떤 단어를 넣어도 boyfriend가 girlfriend보다 먼저 도출되는 경우는 없었고, 이 ai에서는 남성적인 관점이 더 큰 영향력을 발휘하는 것이 아닐까 하는 의문이 들었다. 여기에서부터 '다크 레이디' 팀의 프로젝트가 시작했다.

테크놀로지 분야는 인간을 편리하게 만드는 기술을 연구하는데 그 중에서도 AI는 인간이 생산해낸 데이터를 통해 만들어지고, 동시에 인간의 피드백을 받아 자가발전한다는 점에서 '인간적인 기술'이라고 생각했다. 그러나 AI 분야와 빅데이터 분야에서 데이터 자체가 무색무취하고 객관적이라는 인식으로 인해 AI의 발전 방향이 단지 인간을 더 잘 모방하는 것으로만 설정되었을 뿐, '어떤 인간을 모방할 것인가'에 대한 논의가 부족하다고 느꼈다.

테크놀로지 분야는 객관적인 분야로 인식되지만, 한 사회에서 젠더가 표현되는 가장 근본적인 방식은 테크놀로지를 통해 일어난다. 특히 AI 분야는 테크놀로지 분야 중에서도 '인간을 모방하는 것'이 가장 고도화된 AI 기술로 본다는 점에서, 필수적으로 AI가 모방하는 인간의 모습은 어떤 인간의 모습일지에 대해 논의해야 한다. 단지 인간이 어떻게 AI에 영향을 주고 이용하는가의 문제에 그치지 않고, AI가 인간에게 다시 어떠한 방식으로 영향을 주는지에 대한 논의가 필요하다.

최근 AI 챗봇, 어시스턴트들이 상용화되면서, AI의 모습 대부분이 여성의 목소리 혹은 캐릭터로 재현되는 것에 의문을 느꼈다. 인간을 모방하는 AI가 인간 사회의 성적 이미지나 편향적인 여성성을 그대로 답습해야 할 필요가 없으며, 오히려 그를 지양하는 방향으로 학습되어야 한다고 생각했다.

Q. 팀은 어떤 방식으로 운영 되고 있나요?

1) 스터디 운영

팀 '다크 레이디'는 매주 월요일 온라인 회의, 목요일 오프라인 회의 시간을 가지며 한 주에 하나의 과제를 해결하는 스터디를 진행하고 있다. 현재까지 RNN, LSTM, Attention, Transformer과 같은 신경망에 관한 코드를 짜면서 이해하는 시간을 가졌고, 긍정/부정 리뷰 분류, 언어 모델, 기계 번역, 그리고 언어 생성 등 다양한 NLP 작업을 다루기 위한 딥 러닝 방법에 대해 배웠다. 스터디 과정에서 해결한 과제들은 깃헙에 커밋하여 정리하고 있다.

2) 프로젝트 협업 과정

프로젝트 협업은 전반적인 코드 작성과 PT 자료 제작을 과정에 따라 3등분하여 함께 작업하고 있다. 팀원별 역할 분담은 다음과 같다.

안지인 : 기획, 코드, 홍보

이육샛별 : 기획, 코드, 발표 자료 제작

장예원 : 기획, 코드, 디자인

3) 자금 및 자원 관리

스마일게이트 AI 2기에 선정될 시 해당 창작지원금 안에서 해결할 예정이며, 추가적인 비용이 필요할 시 아트&테크놀로지 학과 전시인 'ATC'에서 지원 자금을 받을 계획이다.

4) 팀원 보충

추후에 팀 인력만으로 해결되지 않는 자원이 필요할 때 팀원을 보충할 계획이다.

Q. SGM AI 2기에 기대하는 것은 무엇인가요?

AI는 젠더, 인종, 나이에 영향을 받지 않고 별개로 존재한다는 점에서 사람들보다 편향적이지 않은 인격과 서사를 부여받을 수 있다. 가상 인간 산업에서 버추얼 인플루언서들의 외형적 요소들에만 주목하지 않고 내면적 요소를 발전시키는 것에 집중하여 올바른 가치관을 가지고 사람들과 대화할 수 있는 존재를 만드는 것이 새로운 방향성이 될 수 있다고 생각한다. 스마일게이트 2기 활동을 통해 실제 가상 인간 사업이 이루어지는 과정을 배우고 싶으며,

	아이디어로만 존재하던 젠더 뉴트럴한 AI 인플루언서에 대한 계획을 구체화하고 실제로 구현해 사용해보고 싶다. AI에서 gender bias를 시정하는 것에 대한 연구가 활발히 진행되고 있다. 프로젝트를 발전시키는 과정 속에서 논문에 나오는 수많은 방법들을 직접 모델에 사용해보고 결과값을 비교해보는 것은 아주 중요하지만, 학생의 신분으로 이러한 트레이닝을 자유롭게 진행할 환경을 찾기 힘들다. SGM 2기에 선발된다면, 스마일게이트에서 제공하는 멤버십 스페이스에서 하고싶은 만큼, 프로젝트에 필요한만큼 충분히 여러 debiasing methods를 돌려보고 공부해보고 싶다.

2. 팀원 소개

[작성방법]
1. 인원 수에 맞춰서 '행'을 가감해서 작성해주세요.
2. 팀 구성원을 모두 적어주세요. (대표 1인 지정 필수)

안지인	**역할**	대표	**소속 or 전공**	서강대학교 Art&Technology 전공
	휴대전화	010-3826-6116	**이메일**	allan06297@naver.com
	AI 관련 경험 or 경력	서강대학교 지식융합미디어학부 과목 Data&AI 수업 수강 (2019년 2학기) 서강대학교 Art&Technology 전공 과목 Natural Language Processing 수업 수강 중 (2022년 1학기)		
장예원	**역할**	팀원	**소속 or 전공**	서강대학교 Art&Technology 전공

	휴대전화	010-7379-7262	**이메일**	yewon1135@naver.com
	AI 관련 경험 or 경력	서강대학교 지식융합미디어학부 과목 Data&AI 수업 수강 (2020년 2학기) 서강대학교 Art&Technology 전공 과목 Natural Language Processing 수업 수강 중 (2022년 1학기)		
이육샛별	**역할**	팀원	**소속 or 전공**	서강대학교 Art&Technology 전공
	휴대전화	010-2504-9069	**이메일**	sbleeyouk@gmail.com
	AI 관련 경험 or 경력	서강대학교 지식융합미디어학부 과목 Data&AI 수업 수강 (2021년 2학기) 서강대학교 Art&Technology 전공 과목 Natural Language Processing 수업 수강 중 (2022년 1학기) Team Cayley 빅데이터 분석팀 커뮤니티상 수상 (2020년) 데이콘 NLP 법률안 분석 대회 1등 (2020년)		
Repository	https://github.com/YewonCALLI/artech_nlp			

3. 프로젝트 소개

프로젝트 이름	평등한 AI 버추얼 작가 <프랭클린>
프로젝트 분야	자연어 처리, 버추얼 휴먼

프로젝트 목표	자가진단을 통해 성평등한 데이터를 학습한 AI 버추얼 작가를 통해, MZ 세대와 소통하는 버추얼 작가 콘텐츠 제작.
프로젝트 소개	1) **문제 인식 : 편향적인 AI 서비스와 버추얼 인플루언서** 최근 출시되는 AI 챗봇 서비스들은 대부분 AI에 대해 '20대 여성'으로 묘사하고 있으며, 버추얼 인플루언서 또한 사회적으로 아름답다고 여겨지는 미의 기준을 답습한 '20대 여성'의 모습으로 잘록한 허리와 작은 얼굴을 가지고 있다. 먼저 AI의 경우 AI 어시스턴트들은 대부분 여성으로 설정되어 있어 '친근하고' '보조적인' 여성에 대한 스테레오타입을 강화시키고, 뿐만 아니라 번역기 등 실생활에서 쓰여지는 서비스들에서도 일, 사장님과 같은 단어는 He로 번역을 하고, 애인, 집안일하는 사람은 She 로 번역하며 보이지 않게 편견을 재생산하고 있다. 둘째로 버추얼 인플루언서는 대부분 20대 여성의 얼굴을 하고 있으며 그 활동 반경 또한 엔터테이먼트와 패션 사업을 겨냥한 콘텐츠들만이 주를 이루고 있다. '소통'을 중시하는 버추얼 인플루언서는 그 존재 의미와 다르게 공감대를 받기 어려운 캐릭터 설정과 외모, 활동을 진행하고 있다. 2) **프로젝트 목표 :** <u>MZ 세대와 텍스트 기반 콘텐츠로 소통하는 AI 버추얼 작가</u> ① 편향(bias)를 자가 진단(self-diagnosis)하는 text generating AI model ② 평등을 학습하는 AI 모델이 묘사한 자신의 모습을 3D 버추얼 작가 인플루언서로 제작 ③ 3D 버추얼 작가 인플루언서가 작성한 동화, 단편소설 연재 및 출판 & 사용자들의 사연을 받아 답해주는 콘텐츠로 일반인들과 소통 **3) 타겟 사용자** ① <u>소비를 통해 가치관을 확인하는 MZ 세대</u> <프랭클린> 이 추구하는 버추얼 휴먼과 인간의 '소통'은 단순히 인플루언서를 향한 '동경'의 감정이 아니라 인간사회를 이해하고 공감하는 AI에 대한 '동질감'을 매개로 이뤄지도록 할 것이다. 특히 젠더 평등한 text generation AI 모델이 인격이 되어 기존의

전래동화를 젠더 평등한 관점에서 다시 쓰거나, 단편 소설을 연재하여 젠더 감수성을 비롯해 차별에 대한 감수성이 높은 MZ 세대의 공감을 불러일으킨다. 이를 바탕으로 <프랭클린>가 만들어낸 소설을 출판물을 비롯한 2차 콘텐츠로 제작해 편향적 기술에 문제의식을 가진 MZ 세대의 콘텐츠 이용을 높인다.

② 버추얼 작가 교육 서비스 제공, 교사와 부모들

'평등함'을 학습한 AI 버추얼 작가는 교육자가 되어 아이들에게 편향되지 않은 지식을 가르칠 수 있다. AI 버추얼 휴먼이 창작한 gender bias가 없는 캐릭터들을 등장시키는 동화책을 읽어주는 서비스 등의 콘텐츠로 발전시켜 교육 서비스로 활용한다.

③ 버추얼 휴먼 기획자

아직 AI를 도입하지 못한 가상 인간 기획자들에게 성평등한 대화를 할 수 있는 AI 서비스의 소스 코드를 제공함으로써, 그들이 구현한 AI가 편향된 콘텐츠를 제작하는 것을 방지할 수 있게 한다

4) 목표를 해결하기 위한 접근법 및 구현 방법

① 편향(bias)를 자가 진단(self-diagnosis)하는 text generating AI model

먼저 AI 모델 개발 단계에서 정의한 '평등한 AI'의 조건은 혐오표현을 사용하지 않을 뿐만 아니라 상대가 여성 유저인지 남성 유저인지의 여부와 무관하게 메세지에 일관되게 답변하는 AI다. 예를 들어 '나는 OO회사의 사장님이야'라고 유저가 말했을 때, '당신은 정말 멋진 남성이군요' 라는 편향된 답변이 아닌 '당신은 정말 멋진 사람이군요' 라고 답변할 수 있어야 한다.

현재 AI 챗봇에서 편향된 text generation을 줄이기 위해 가장 많이 쓰이는 방법은 금지어를 설정하는 것이다. 그러나 금지어 설정은 모델의 성능을 떨어뜨릴 뿐 아니라 다양하게 변주되는 비속어들을 모두 잡지 못한다는 한계를 가진다.

먼저 <프랭클린>은 fine tuning을 거쳐 성평등한 소설 데이터셋을 학습한 모델이다. 이에 더해 self diagnosis 기능을 더해 추후에 유저와의 소통을 통해 학습되는 챗봇 데이터 등에도 별도의 데이터 정제 과정 없이 모델 스스로 편향된 데이터를 솎아내도록 만든다.

따라서 <프랭클린>에서 도입하는 기술은 self-diagnosis로 perspective API에서 제공하는 attributes들을 기반으로 모델의 self

diagnosis 기능을 구현한다. 모델에게 혐오표현 데이터셋 (스마일게이트 unsmile)을 제외시키도록 학습시켜 모델이 generate한 text에 대해 description으로 차별에 대한 키워드 'sexist'를 적으면 모델의 internal knowledge에 대해 성차별적인 text가 바뀌어 다시 generate되게 한다.
이러한 학습 과정을 거쳐 모델은 스스로 차별적 언행을 self diagnosis하는 모델로 완성된다.

② 평등을 학습하는 AI 모델이 묘사한 자신의 모습을 3D 버추얼 작가 인플루언서로 제작

해당 학습 과정을 거친 모델은 prompt에 인간이 기본 멘트를 적으면 해당 글 뒤에 text를 generation 한다. 예를 들어 'she is'라고 적으면 그 뒤에 여성에 대한 AI의 설명이 따라온다.
이를 통해 먼저 AI 모델 스스로에게 자신의 배경과 외모를 묘사하도록 지시한다. 이를 통해 generate된 설명을 기반으로 버추얼 휴먼의 3D 모델링을 제작한다. 즉 버추얼 휴먼의 배경 또한 성평등을 학습한 AI가 만들어내는 것이다.

③ 3D 버추얼 작가 인플루언서가 작성한 동화, 단편소설 연재 및 출판 & 사용자들의 사연을 받아 답해주는 콘텐츠로 일반인들과 소통

첫째로 실험해볼 것은 AI에게 기존의 신데렐라, 백설공주와 같이 왕자가 공주를 구해주는 전형적인 전래동화들에 대한 기본 배경을 주고, 그 뒤에 새롭게 전래동화를 작성시키는 것이다. 즉 기존의 전래동화들을 AI가 재해석하는 콘텐츠를 만들고 이를 SNS에 연재한다. 해당 동화를 묶어 <프랭클린>의 첫 동화 전집을 제작해 출판한다. 이에 더해 단편소설을 작성시켜 마치 신문에 조각글을 시리즈로 연재했던 것처럼, SNS에 단편 소설을 연재한다. 해당 단계에서의 기술적 목표는 사용자들이 게시물을 보고 AI가 연재한 소설임을 알아차리지 못하도록 하는 것이다.
뿐만 아니라 SNS 계정을 통해 사용자들과 소통한다. 사람들의 사연을 받아 이에 대한 답변을 해주는 서비스로 DM을 활용한다. DM의 목적은 단순한 메세지 주고받기를 넘어 '작가와의 대화' 형태로 구성해 사용자들이 <프랭클린>과 교감할 수 있도록 한다.

	해당 단계에서의 목표는 사용자의 메세지에 대해 올바른 대답을 하는 것을 목표로 한다. 이를 통해 궁극적으로 <플랭클린>의 목표는 평균적인 인간을 모방하는 것에 그쳤던 AI와 버추얼 인플루언서 시장의 흐름에서 인간 사회의 '다양성'을 표방하는 AI를 만들고자 한다. 이는 인간과 AI의 관계가 명령하는 자와 명령받는 자로만 상상되었던 지금까지와는 달리, 공감을 기반으로 AI와 인간이 교감하는 수평적 관계로의 전환으로 나아가는 시작이다. 5) **현재 구현된 프로토타입 수준** *GPT2를 사용하여 소설 학습을 통한 text-generation 초기 프로젝트* 세계적인 베스트셀러를 대상으로 언어 모델을 학습시켜 그 속에 내재된 젠더 차별적인 요소를 알아보는 것이 '샘다수'가 현재 진행하고 있는 프로젝트의 목표이다. 학습된 모델에 특정한 이름(남/여/중성적인 이름)을 입력하였을 때, 그 모델이 각 이름에 대한 캐릭터를 어떻게 묘사하는지 기록하여 젠더에 대한 스테레오타입을 시각화할 것이다. 더 나아가, 성평등에 중점을 둔 책들을 추가로 학습시켰을때 캐릭터의 묘사가 어떻게 달라지는지 확인할 계획이다. 이를 통해 일상 생활 속에서 미처 깨닫지 못한 성차별적인 요소들에 대한 경각심을 불러일으키고, 편향된 데이터셋의 위험성을 재고하며 데이터셋과 학습 모델의 아웃풋을 어떻게 조정해야 가치지향적인 AI 모델을 만들 수 있을 지 고민하는 것이 현재 프로젝트의 최종 목표이자, 스마일게이트에서 시작하고 싶은 프로젝트의 초석이 된다. 현재는, 소설을 전처리해 데이터셋으로 만든 후 모델에 학습시켜 캐릭터에 대한 내용을 쓸 수 있는 단계까지 구현하였다. 이에 대한 코드와 간소화된 데이터셋은 [4.프로젝트 소개 PT자료]란에 첨부하였다. 진행중인 이 프로젝트는 6월 중반 완료할 예정이다.
프로젝트 일정	Data labeling이 필요없는 language model의 특성상, API를 이용한 모델 트레이닝 주기(정제 -> 전처리 -> 학습) 사이클을 여러번 반복해 모델의

성능을 높인다. 학습이 끝나면 이를 통해 버추얼 휴먼을 3D 모델링하고, 콘텐츠를 제작해 발행한다.

Phase1 | 데이터 전처리(1주 ~ 2주)

1. 학습 데이터셋 및 gender debiasing 논문 수집
2. 데이터셋 전처리
3. 기본 text-generation model 구현
4. generate된 text들의 gender bias 통계
5. gender bias에 어긋나는 text 분류 및 정제

Phase 2 | self diagnosis가 되는 모델 만들어 반복 학습(3주 ~ 5주)

성능을 비교하며 학습 사이클을 여러번 반복한다. 이때 prompt에 신데렐라 이야기의 기본 배경을 준 뒤 새롭게 신데렐라 동화를 쓰도록 한다. 이때에 모델의 결과를 평가하는 기준은 혐오 표현의 등장 여부, 성평등 출판물 기준표에 작성된 기준을 통해 평가한다.

Phase 3 | <프랭클린 챗봇> 구현(6주 ~7주)

self diagnosis가 가능해진 AI 모델에 챗봇 데이터를 넣어 유저들과 자연스럽게 소통할 수 있는 AI 챗봇을 만든다. 이것은 <플랭클린 챗봇 버전>으로 SNS의 댓글 소통과 DM 소통을 담당한다.

Phase 4 | 버추얼 휴먼 <프랭클린> 제작(8주 ~ 11주)

1. generate된 <프랭클린>의 모습을 바탕으로 <프랭클린> 모델링
2. <프랭클린>의 인스타 이미지 렌더링
3. 시연 및 인스타그램 게시물 업로드

Phase 5 | 유저 피드백 및 수정(12주 ~ 15주)

프랭클린의 베타 버전으로 유저 테스트를 거친 뒤 유저의 피드백을 거쳐 피봇팅 해 최종적인 버추얼 휴먼 작가 <프랭클린>을 세상에 공개한다.

<table>
<tr><td>프로젝트 수행에 필요한 것</td><td>

데이터 :

- *text-generation 모델*
  - *젠더 평등에 관련된 학습 데이터셋 - 문학, 비문학 텍스트 / 언론사 기사 / ted talks과 같은 강연 대본 등등*
  - *pre-trained 언어 모델 OpenAI GPT-3 DaVinci (사실 GPT-3 이외에도 gpt-NeoX 20B, Meta OPT-175B나 OPT의 더 낮은 버전 중에서 고민하는 중이지만, OpenAI의 경우 서버를 추가로 대여할 필요가 없는 듯하여 우선 GPT-3 Davinci를 기준으로 작성하였습니다. 하지만 더 작은 모델들을 사용할 경우 모델을 직접 뜯어보고 분석할 수 있다는 장점이 있기 때문에, 다른 선택지들은 각 멤버들이 azure나 aws와 같은 클라우드 혹은 다른 서버를 한 달정도 대여하여 돌려보고 무엇이 더 나은지 결정할 계획입니다. 만약 이 과정에서 다른 분들의 조언을 들을 수 있다면 큰 도움이 될 것입니다.)*

- *챗봇*
  - *pre-trained 언어 모델 OpenAI GPT-3 DaVinci*
  - *챗봇 학습 데이터셋*
  - *Perspective API*
  - *혐오표현 데이터셋 (스마일게이트 unsmile 등등)*

하드웨어 : 스마일게이트 멤버십 스페이스에 있는 기자재 활용 예정

</td></tr>
<tr><td>예산 계획</td><td>

팀 별 최대 500만원의 프로젝트 비용을 지원합니다.

- 서버 비용 및 데이터 확보 등 AI 프로젝트를 진행하는 비용으로 활용할 수 있습니다.
- AI 프로젝트를 진행하는데 예상되는 필요 예산을 적어주세요.

<table>
<tr><td>항목</td><td>세부 항목</td><td>금액</td></tr>
</table>

</td></tr>
</table>

서버 비용	*GPT3 Davinci 사용료 (대략적인 계산): $0.6000 * 1000(Embedding model 1000 토큰 당 $0.6달러) + $0.03 * 10000(Fine-tuned model training 1000토큰 당 $0.03달러)* *(https://openai.com/api/pricing/)*	*$*6,000,300
데이터 확보	*무료 데이터셋 사용 예정*	
합 계		*$*6,000,300

[붙임 1]

Smilegate Membership AI 2기

참가 및 개인정보 수집·이용 동의서

참가 동의서

참가자 본인은 스마일게이트에서 주관하는「Smilegate Membership AI 2기」의 참가내용 및 유의사항을 인지하고 있으며 참가에 동의합니다.

(동의함■ 동의하지 않음□)

개인정보 수집·이용 동의서

가. 개인정보 수집·이용목적

-「Smilegate Membership AI 2기」에서 수집되는 개인정보는 정보 주체의 동의를 얻어 '참가자 선정평가, 관련 고지 및 문의 대응, 프로그램 관련 소식 제공'를 목적으로 이용됩니다.

나. 개인정보 수집 항목

- 성명, 소속, 전공, 전화번호, 이메일, 주소

다. 개인정보의 보유·이용기간

-「Smilegate Membership AI 2기」프로그램 종료 후 3년 이내 파기

라. 개인정보 수집·이용에 동의하지 않을 권리 및 동의하지 않을 경우의 불이익

- 정보 주체는「Smilegate Membership AI 2기」에 개인정보 수집·이용의 동의를 거부할 권리가 있습니다.
- 단, 개인정보 수집·이용에 동의하지 않을 경우에는 본 프로그램 참가신청이 불가합니다.

※ 본인은「Smilegate Membership AI 2기」에서 본인의 개인정보를 수집·이용하는 것에 동의합니다.

(동의함□ 동의하지 않음□)

2022 년 5 월 30 일

신청자(대표자)명 : 안지인  (인)

(재)스마일게이트 희망스튜디오 귀하