

데이터 수집 및 저장 프로젝트 기획서

1. 개요

- 1) 산출물 단계 : 데이터 수집 및 저장
- 2) 평가 산출물 : 프로젝트 기획서
- 3) 제출 일자 : 2025.02.06
- 4) 깃허브경로: <https://github.com/SKNETWORKS-FAMILY-AICAMP/SKN06-FINAL-2Team>
- 5) 작성 팀원 : 박미현

2. 프로젝트 주제

- 웹툰·웹소설 대화형 추천 서비스

3. 문제 정의

3.1 사용자 경험의 문제점, 추천 시스템 한계

- 웹툰·웹소설 콘텐츠 데이터의 폭발적인 증가, 사용자들이 자신의 취향에 맞는 콘텐츠를 찾기 어려움
- 기존의 추천 시스템은 사용자의 세부적인 요구사항을 반영하기 어려움
- 사용자의 취향 변화나 상황에 따른 실시간 추천이 부족함
- 새로운 장르나 작품에 대한 사용자의 탐험을 장려하는 추천 부족
- 웹툰 이용시 플랫폼에서 제공하는 추천 서비스가 빈약하거나 충분하지 못함
- 최초 사용자 조사 이후, '설문 응답자 : 이용 비율'에서 미미하지만 응답 비율이 줄어드는 것은 'Ⅰ.서비스 플랫폼의 다양화, Ⅱ.사용자의 이탈, Ⅲ.불법 서비스 제공하는 사이트가 증가' 등이 원인이 될 것이라 예측
- 기존 서비스 플랫폼에서도 차별화된 추천 서비스로 기존 고객의 이탈을 줄이는데 매력적인 서비스가 필요해질거라 예상

그림 2-1-4 만화웹툰 평균 이용 비율

(Base: 웹툰 및 출판만화 동시 이용자, 평균(%))

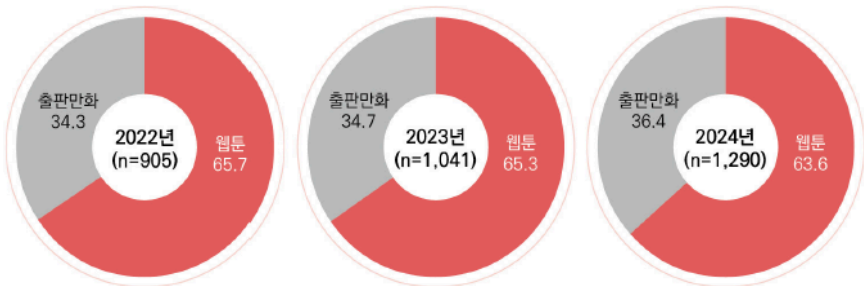


표 2-2-16 웹툰 이용 시 고려 기준 - 작품 외부 요소(1순위)

(Base: 웹툰 이용자, 단위: %)

구분	사례 수	인기순	가격 (유/무료)	최신작	원작이 되는 웹소설에 대한 흥미	자인/만화 커뮤니티 의 추천	쉬운, 편리한 구입/이용	웹툰을 원작으로 하는 영화나 드라마, 웹소설 등 (2차적 저작물)의 흥행	댓글의 평가	서비스 내 추천작
전체	(3,448)	36.4	15.7	5.0	3.5	3.5	2.3	2.0	1.3	1.2
성	남성	(1,798)	41.4	15.7	6.7	3.1	3.2	1.2	1.3	1.1
	여성	(1,650)	31.1	15.6	3.2	4.0	3.8	2.2	1.4	1.3
연령	10대	(458)	42.8	10.6	2.8	2.1	3.8	1.1	1.3	0.3
	20대	(650)	39.6	10.8	3.6	3.0	2.3	0.6	1.2	0.8
	30대	(665)	38.9	11.6	2.8	1.3	4.7	1.3	0.4	0.5
	40대	(642)	36.7	14.8	4.7	4.3	3.1	2.3	1.8	1.7
	50대	(584)	30.9	20.5	6.9	6.1	3.3	2.8	1.7	2.3
	60대	(449)	28.5	28.9	10.2	4.7	3.5	5.7	1.7	1.9

3.2 웹 추천 서비스에서의 할루시네이션 오류

- ChatGPT가 제공하는 정보에 할루시네이션이 존재(네이버-완결 웹툰을 카카오편에서 연재중인 것으로 표현하는 등의 부정확한 정보 전달)

NAVER 웹툰 | 웹소설 | 시리즈

홈 웹툰 베스트도전 도전만화 마이페이지



유미의 세포들

이동건 · 글/그림 | 512화 완결 · 12세 이용가

유미는 지금 무슨 생각을 하고 있을까?
그녀의 머릿속에서 바쁘게 움직이는 세포들 이야기!

#로맨스 #명작 #드라마&영화 원작웹툰 #완결로맨스

*유미의 세포들*에 대해 더 깊이 알려줄게! 😊

기본 정보

- **작가**: Donggeon
- **장르**: 로맨스, 코미디, 일상
- **연재 플랫폼**: 카카오편에서 연재되고 있어!

줄거리

이야기는 평범한 사무직 여성 '유미'가 주인공이야. 그녀의 머리 속에는 다양한 세포들이 살고 있는데, 이 세포들은 감정

주요 캐릭터

- **유미**: 주인공으로 직장 생활에 고군분투하며 세포들과의 다양한 에피소드를 겪어.
- **세포들**: 유미의 감정을 책임지고 있는 다양한 세포들로, 사랑 세포, 먹세포, 열정 세포 등 다양한 역할을 가지고
- **남자 주인공**: 유미와의 사랑 이야기를 이끌어가는 인물로, 둘의 관계 발전이 큰 재미 요소야.

독특한 점

3. 문제 정의

4. 시장 조사

4.1 국내 및 글로벌 웹툰·웹소설 시장 규모 분석

- 국내 웹툰산업 총 매출액 2022년 기준 1조 8,290억 원(전년 대비 16.9% 증가)
- 글로벌 웹툰 시장 규모 : 연평균 36.8% 성장 (출처: [경기콘텐츠진흥원](#))
- 11개 플랫폼의 매출을 모두 합하면 약 1조 3,655억 원이며, 100%의 트래픽으로 환산하면 약 1조 4,094억 원(1,409,458,057,878원)으로 추정됨. 2022년 추정값인 1조 1,277억 원(1,127,731,457,215원) 대비 약 25.0% 증가
- 본 조사의 추정 과정을 통해 산출된 웹툰 CP사의 2023년 총 매출 추정금액은 약 7,795억 원(779,498,626,754원)으로 추정됨
- 웹툰산업의 전체 규모 추정 결과, 2023년 웹툰산업의 매출 규모는 약 2조 1,890억 원 (2,188,956,684,632원)으로 추정됨
- 2019년에서 2020년의 산업 규모 상승률은 65.3%, 2020년에서 2021년은 48.4%, 2021년에서 2022년은 16.8%, 2022년에서 2023년은 19.7%로 전년도
- 상승률보다는 소폭 증가한 수준으로 나타남
(출처: [2024 웹툰산업 실태조사](#))
- "美, 웹툰 산업을 공인된 산업으로 인정"
(출처 : <https://news.einfomax.co.kr/news/articleView.html?idxno=4315035>)

4.2 웹툰 추천 현황

1. Community (한국 - 디시인사이드, 외국 - 레딧)
 - 질문글
 - 추천글
2. Youtube유튜브
3. Web
 - 블로그
 - 개인 홈페이지

4.3 웹툰 가이드 ("<https://www.webtoonguide.com/ko/>")

1. 다양한 플랫폼의 최근 5년간 웹툰·웹소설 업로드됨
2. 플랫폼별 만화, 에이전시별 만화를 모아서 볼 수 있음
 - 플랫폼의 경우 복지 현황, 투고 방법을 명시
 - 에이전시의 경우 원고 투고 방법을 명시
3. 공모전 리스트, 외주계약 시 주의사항, 예술인 복지제도 등 웹툰·웹소설 작가 및 지망생을 위한 정보 제공
4. 소속 직원으로 보이는 관리자와 소속 기자의 정기적인 콘텐츠 업로드
 - 작가 인터뷰 진행
 - 추천글 작성(매일 혹은 이틀에 한 번 정도 비정기적인 주기)
 - 카드뉴스 제작
 - 웹툰·웹소설 관련 기사 작성
 - 직원 및 한국만화영상진흥원 작성 컬럼 존재(2020.12.20 마지막 업로드)
 - 영어 인터뷰, 리뷰, 컬럼 존재(2019~2021 마지막 업로드)

<p>4. 시장 조사</p>	<ul style="list-style-type: none"> 5. 콘텐츠 데이터 & AI를 다루는 코니스트社에서 운영하는 홈페이지 <ul style="list-style-type: none"> ○ 자체 서비스 중인 ‘웹툰·웹소설’이 상업 데이터베이스 서비스를 토대로 운영 6. 대중화된 플랫폼보다 가독성이 낮은 UI/UX 디자인으로 사용자 이용 호감도 저하 예상 7. 광고가 빈번하게 팝업되어 사용자 불편 초래 8. 성인물 썸네일 작품 다수 업로드되어 있으나 인증 절차 없음 <p>4.4 GetGPT ("https://getgpt.app/play/RjUr2RnLFX")</p> <ul style="list-style-type: none"> 1. AI 제작 서비스 플랫폼에서 생성된 개인 창작물 2. 선호하는 장르와 플랫폼을 입력하면 해당 플랫폼의 웹툰을 추천하는 답변 생성 (약 1392자) 3. 답변 언어는 영어만 존재하여 한글 사용자에게 불편 예상 4. 데이터셋 업데이트 지연(최소 1년 전 업로드된 것으로 확인) 5. 무료 6. 광고 없음 <p>4.5 Talkie ("https://www.talkie-ai.com/chat/recommend-webtoon-140666908045394")</p> <ul style="list-style-type: none"> 1. Talkie AI 제작 서비스 플랫폼에서 생성된 개인 창작물 2. ChatGPT와 유사한 디자인의 채팅 화면 3. 디자인 가독성이 떨어짐 4. 서비스 검색 시 광고 팝업 발생 5. TTS(Text to Speech) 기능 제공 6. 장르나 웹툰 종류 입력 시 추천 가능 7. 전구 모양 아이콘 클릭 시 추천 질문 제공 8. 이미지 검색 버튼 제공 (사이트 자체 기능) <p>4.6 사용자 조사 및 설문 결과</p> <p>선호하는 장르, 작가를 기반으로 한 개인화 추천에 대한 관심 파악. 대화형 추천 서비스에 대한 사용 편의성과 매력도 등의 다양한 설문 및 작성된 기사 확인. (" https://www.seoulwire.com/news/articleView.html?idxno=604713 ")</p>
<p>5. 모델링 방안</p>	<p>5.1 자연어 처리(NLP) 기반 챗봇 : 사용자가 입력한 대화형 질문에 따라 취향 정보를 파악하고 작품을 추천.</p> <p>5.2 RAG 챗봇(웹툰·웹소설 데이터 이용) : 대표 플랫폼인 네이버, 카카오에서 제공하는 웹툰·웹소설의 정보를 임베딩 벡터로 저장하여 챗봇이 정확한 추천 제공.</p> <p>5.3 협업 필터링 및 콘텐츠 기반 필터링 혼합 모델: 유사한 취향의 사용자가 좋아하는 작품을 분석하고, 작품의 속성을 기반으로 추천 제공.</p>

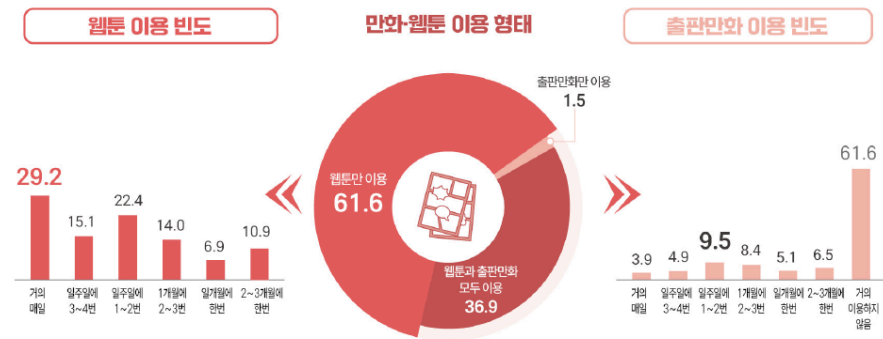
<p>6. 사용데이터</p>	<p>6.1 웹툰·웹소설 메타데이터: 대표 웹툰·웹소설 플랫폼 크롤링</p> <ul style="list-style-type: none"> • 네이버 웹툰 • 네이버 시리즈 웹소설 • 카카오웹툰 • 카카오페이지 웹툰 • 카카오페이지 웹소설 • 플랫폼, 종류, 제목, 연재 상태, 썸네일, 장르, 조회수, 별점, 좋아요, 줄거리, 키워드, 글 저자, 그림 저자, 원작자, 연령 제한, 가격, url, 총 회차 수 <p>6.2 사용자 취향 데이터</p> <ul style="list-style-type: none"> • 회원 가입시 사용자의 취향 파악 • 썸네일을 보고 읽어봤거나 읽어보고 싶은 작품 선택 • 작품: 웹툰·웹소설 각각 32개 • 선택한 작품의 키워드 등의 정보로 최초 취향 구성 • 챗봇을 통해 추천한 작품에 대한 피드백
<p>7. R&R</p>	<p>7.1 백엔드</p> <ol style="list-style-type: none"> 1) 챗봇 모델 <ul style="list-style-type: none"> • 로판 챗봇 담당 (전하연) • 로맨스 챗봇 담당 (정유진) • 히어로 챗봇 담당 (안형진) • 시대물/무협 챗봇 담당 (박미현) • 기본 챗봇 담당 (이세화) 2) django 웹 <ul style="list-style-type: none"> • 애니메이션 및 사이트 로고 담당 (3명) • 메인 화면 담당 (2명) <p>7.2 데이터</p> <ul style="list-style-type: none"> • 웹툰·웹소설 meta data, sample data (정유진) • 사용자 users data(안형진) • 사용자 피드백 data(박미현,이세화) <p>7.3 산출물</p> <ul style="list-style-type: none"> • 화면설계서(전하연) • 프로젝트기획서(박미현) • 회의록(정유진)

8. 만화-웹툰 이용자 조사

“2023년 6월 ~ 2024년 5월 (최근 1년) 간의 전국 만 10~69세 만화-웹툰 콘텐츠 이용자를 대상으로 조사한 한국콘텐츠진흥원 에서 작성한 자료를 참고하였다.”

(“ <https://welcon.kocca.kr/ko/info/trend/1954561> ”)

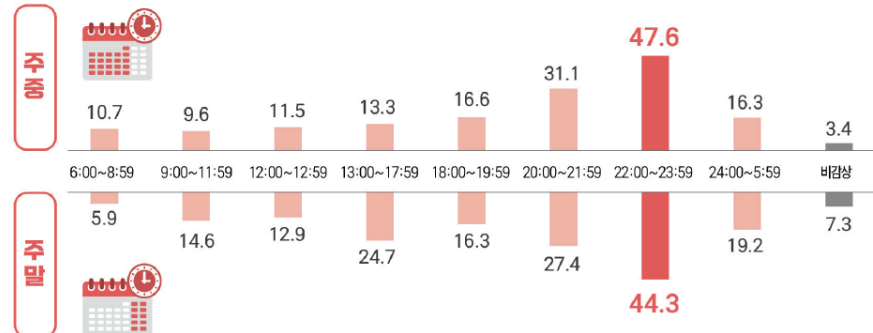
8.1 만화-웹툰 이용 형태 및 빈도



8.2 웹툰 만화 이용 현황

1) 감상 시간대

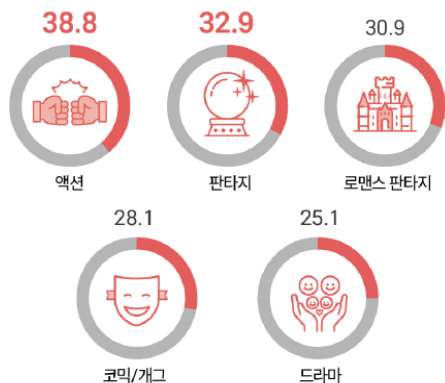
1 웹툰 감상 시간대 (Base : 웹툰 이용자, 중복응답, 단위 : %)



2) 즐겨보는 장르

2 즐겨보는 웹툰 장르 TOP5

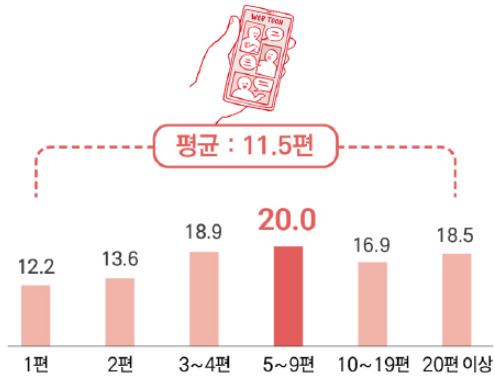
(Base: 웹툰 이용자, 단위 : %, (1+2+3순위))



3) 작품 감상 수

3 주당 웹툰 작품(회차 기준) 감상 수

(Base: 웹툰 이용자, 단위: %)



4) 주요 이용 플랫폼

4 온라인만화 주 이용 서비스 (Base: 온라인만화 이용자, 중복응답, 단위: %)

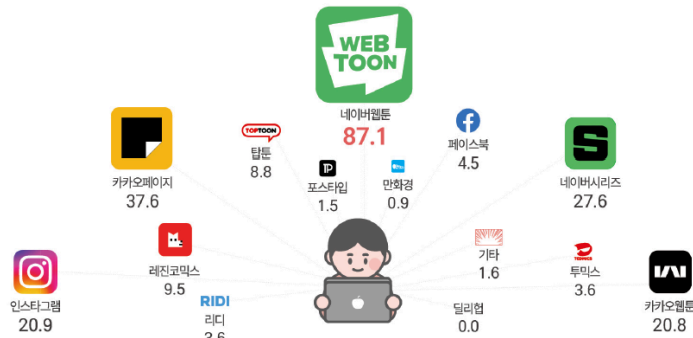
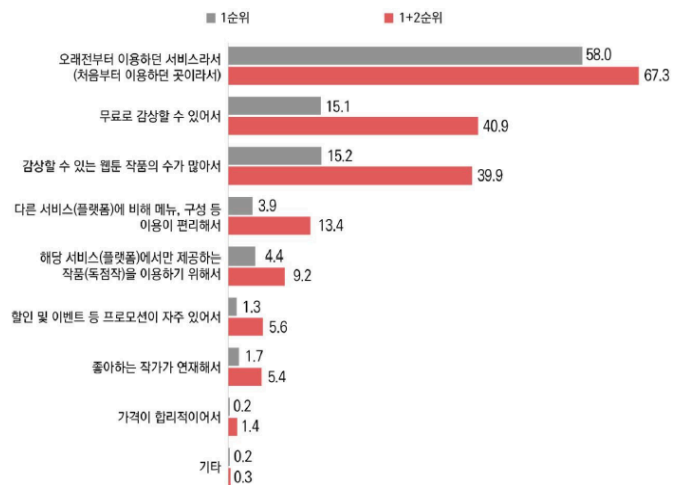


그림 2-2-8 온라인만화 주 이용 서비스 이용 이유

(Base: 온라인만화 이용자, 단위: %)



주 1) 2024년 '좋아하는 작가가 연재해서' 보기 추가
2) 2024년 '가격이 합리적이어서' 보기 추가

5) 이용자들이 즐겨보는 작품에 대한 설문

표 2-2-13 즐겨보는 웹툰 작품 유무

(Base: 웹툰 이용자, 단위: %)

구분	사례 수	있음	없음
전체	(3,448)	73.3	26.7
성	남성	75.6	24.4
	여성	70.8	29.2
연령	10대	86.6	13.4
	20대	88.7	11.3
	30대	85.3	14.7
	40대	74.1	25.9
	50대	55.8	44.2
	60대	41.2	58.8

6-1) 이용자가 작품을 선정하는 기준

표 2-2-17 웹툰 이용 시 고려 기준 - 작품 외부 요소(1+2+3순위)

(Base: 웹툰 이용자, 단위: %, (1+2+3순위 기준))

구분	사례 수	인기순	가격 (유/무료)	최신작	지인/만화 커뮤니티 의 추천	원작이 되는 웹소설에 대한 흥미	댓글의 평가	쉬운, 편리한 구입/이용	웹툰을 원작으로 하는 영화나 드라마, 웹소설 등 (2차적 저작물)의 흥행	서비스 내 추천작
전체	(3,448)	47.8	31.0	14.5	10.3	8.9	8.3	7.8	7.3	4.4
성	남성	51.6	31.5	19.2	10.4	8.2	8.8	8.7	6.4	4.3
	여성	43.7	30.4	9.4	10.1	9.7	7.7	6.7	8.2	4.4
연령	10대	50.9	26.4	11.4	9.0	7.0	7.7	4.6	6.1	4.2
	20대	50.3	24.0	14.4	9.7	6.7	7.3	3.3	1.6	3.2
	30대	50.0	23.8	10.9	11.2	5.4	6.6	6.1	2.4	2.8
	40대	47.5	29.5	15.4	9.3	10.4	9.0	6.8	8.9	4.1
	50대	46.0	39.2	17.4	10.8	12.4	9.6	10.9	13.7	5.6
	60대	40.9	47.9	18.2	11.7	12.5	9.9	17.2	13.2	7.3

6-2) 이용자가 작품을 선정하는 기준

표 2-2-19 웹툰 이용 시 고려 기준 - 작품 내부 요소 및 기타(1+2+3순위)

(Base: 웹툰 이용자, 단위: %, (1+2+3순위 기준))

구분	사례 수	소재/ 줄거리	그림/ 그림체	장르	작품성/ 예술성	특정 작가의 작품	완결 여부	기타
전체	(3,448)	39.9	34.7	34.4	12.9	5.7	5.0	0.3
성	남성	36.2	27.3	35.2	13.0	4.9	3.9	0.3
	여성	44.0	42.9	33.6	12.8	6.5	6.2	0.3
연령	10대	36.4	47.2	42.5	14.6	3.4	4.6	-
	20대	43.7	50.7	36.9	15.5	6.5	2.7	0.2
	30대	47.9	43.4	40.3	13.6	6.9	4.5	0.9
	40대	37.6	32.7	28.8	13.7	8.0	6.2	0.2
	50대	35.6	19.0	30.3	9.0	3.5	7.4	0.3
	60대	35.0	9.4	27.4	10.3	4.5	4.6	0.1

7) 이용자가 웹툰 정보는 획득하는 경로

표 2-2-2 웹툰 정보 획득 경로(1+2+3순위)

(Base: 웹툰 이용자, 단위: %, (1+2+3순위 기준))

구분		사례 수	포털사이트	친구나 주변인	SNS	인터넷 커뮤니티/ 카페/갤러리	유튜브
전체		(3,448)	57.0	38.5	34.2	30.1	30.0
성	남성	(1,798)	57.1	37.1	26.7	34.9	33.2
	여성	(1,650)	57.0	39.9	42.4	24.9	26.5
연령	10대	(458)	29.1	59.4	51.6	19.7	47.0
	20대	(650)	46.5	45.1	48.7	24.7	35.5
	30대	(665)	61.5	38.6	34.6	26.3	21.9
	40대	(642)	64.4	31.8	25.6	35.2	23.5
	50대	(584)	69.2	30.2	20.9	36.5	25.0
	60대	(449)	67.8	27.6	24.6	38.6	32.2

구분		사례 수	웹툰 전문 사이트	웹툰 원작 2차적 저작물	자상파TV/ 신문/잡지/ 인터넷의 기사	만화 카페	단행본에 들어있는 광고	기타
전체		(3,448)	22.8	18.3	11.6	6.5	4.2	0.6
성	남성	(1,798)	26.2	12.8	12.9	8.3	4.8	0.4
	여성	(1,650)	19.2	24.3	10.2	4.5	3.5	0.9
연령	10대	(458)	21.8	14.7	6.9	9.2	5.7	0.3
	20대	(650)	22.1	18.8	4.5	6.4	4.3	1.3
	30대	(665)	18.7	23.5	7.0	3.0	3.7	1.1
	40대	(642)	23.8	20.0	12.3	7.2	3.4	0.3
	50대	(584)	25.3	18.4	17.3	6.8	4.2	0.3
	60대	(449)	26.5	11.2	25.1	7.7	4.3	0.3

8) 그외, 웹툰이 높은 이용률을 보일 수 있는 이유는 학생, 직장인 등 바쁜 일상 속에서 짧은 시간에 소비할 수 있는 콘텐츠이기 때문이며, 이를 통해 스트레스 해소와 소소한 행복을 느낀다는 의견이 도출되었다.

9) 포털사이트의 웹툰 플랫폼이 높은 이유는 구독 편의성과 접근성이 용이하며, 장르가 다양하기 때문이었다. 또한, 공급의 증가로 인하여 높아진 웹툰의 품질(퀄리티)과 요일별 신규 회차를 감상할 수 있다는 점에서 큰 강점이 있는 것으로 나타났다.

! 반면, 대형 플랫폼에서는 추천 서비스가 존재하지만, AI와의 채팅을 통해 추천받는 서비스는 없었다. 이용자는 원하는 취향의 콘텐츠를 얻기 위해 단순하고 반복적인 여러 단계를 거쳐 스스로 선택해야 한다.

방대한 양의 모든 작품에서 직접 줄거리를 확인해야 하며, 다른 이용자들의 피드백이 각기 달라 타인의 취향을 참고하기 어렵다. 또한, 이용하고자 하는 콘텐츠의 리스트를 만들어 관리해야 하는 번거로움이 있다.

!! 우리가 구현하려는 추천 시스템을 통해 이용자는 원하는 취향과 장르의 콘텐츠를 보다 쉽게 찾고, 리스트로 관리할 수 있도록 하는 것을 목표로 한다.