



# 글:림 (GL:IM)

## AI기반 독서 리딩 메이트

글(Text)을 그림(Image)으로, 아이의 이해를 '빛(Gleam)'으로 비추다 6~9세 언어 발달 장애 아동 & 부모 "One Story, Infinite Worlds"

**VREW**



# 팀 소개 - 육안(肉眼)

육안(肉眼): 6개의 눈

"글자만으로는 세상을 상상하기 힘든 아이들이 있습니다.  
저희 팀- '육안(肉眼)'은 글자 뒤에 숨겨진 의미를 아이들이  
**자신의 눈(육안)으로 직접 보듯** 선명하게 느낄 수 있어야 한다고  
생각합니다."

# 팀 소개 - 육안(肉眼)

6명의 따뜻한 시선으로  
바라보는 우리 아이들의 세상

## 팀의 정체성

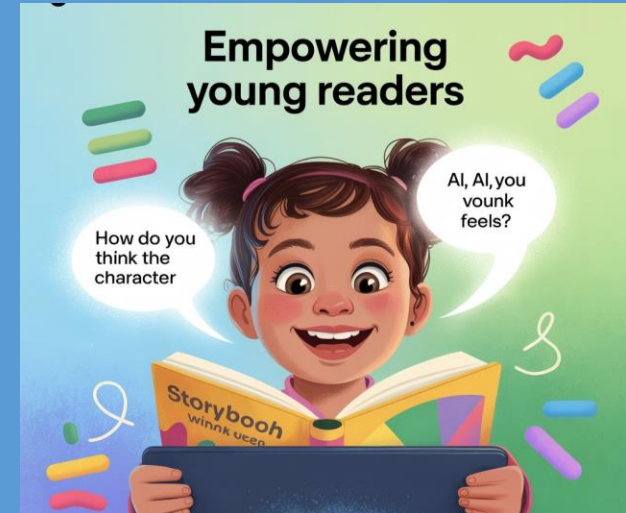
육안(肉眼)은 6개의 시각(6명의 팀원)이 모여 따뜻한 시선으로 세상을 바라봅니다. 저희는 글자 뒤에 숨겨진 의미를 아이들이 자신의 눈(육안)으로 직접 보듯 선명하게 느낄 수 있도록 돕고자 합니다.

## 미션 & 비전

저희는 텍스트 뒤에 숨겨진 이야기를 시각적으로 연결해 주는 가교가 되겠습니다.  
아이 곁에서 즐거운 Reading(독서)을 돕고, 궁극적으로는 언어 능력을 Leading(향상)시키는 든든한 페이스메이커가 되고 싶습니다.

## 핵심 가치

근거 기반(Evidence-based): 언어발달 장애 아동 연구에 기반한 솔루션 설계  
포용 성(Inclusivity): 다양한 학습 방식을 지원하는 맞춤형 접근  
안전한 AI(Safe AI): 아동 친화적이고 안전한 AI 활용  
확장 성(Scalability): 지속적인 발전과 다양한 환경 적용 가능



# Empowering young readers

## 서비스 소개 글:림(GL:IM)

How do you  
think the  
character

AI, AI, you  
vounk  
feels?

글:림은 '글(Text)'을 '그림(Image)'으로 변환하고, AI 질문 유도를 통해 아동의 자발적인 감정 표현과 상호작용을 촉진하는 서비스입니다. 텍스트의 장벽에 가로막힌 아이들에게 독서의 즐거움을 비추는 작은 '빛(Gleam)'이 되고자 합니다.

### AI 생성 시각화

텍스트를 분석하여 즉시 고유한 이미지를 생성합니다. 아이의 이해를 돕고 상상력을 자극하며, 매번 다른 시각적 해석으로 N회차 독서를 가능하게 합니다.

### 맞춤형 대화 유도

생성된 이미지에 기반한 맞춤형 질문을 통해 아이의 자발적인 감정 표현 및 대화를 유도합니다. 부모와 아이의 의미 있는 상호작용을 촉진합니다.

### 학습 보조 게임

Custom Vision(Object Detection)을 활용한 숨은그림찾기 미니 게임을 통해 아이의 인지적 이해 및 집중력을 강화합니다. 놀이를 통한 자연스러운 학습 경험을 제공합니다.

### 대상 사용자

6~9세 언어 발달 장애 아동과 부모를 주요 대상으로 합니다. 특히 시각적 정보 처리가 언어보다 강점인 아동(ASD, DLD 포함)과 함께하는 부모님들에게 효과적인 독서 보조 도구입니다.

# 목차

# CONTENTS

## 1.프로젝트 개요

팀소개, 서비스 소개, 시장규모, 경쟁사분석, 시장 포지셔닝맵

## 2.프로젝트 팀 구성 및 역할

언어 발달 장애 아동의 독서 어려움, 타겟 사용자

## 3.프로젝트 수행 절차 및 방법

시스템아키텍처

## 4.프로젝트 수행 경과

Ui, UX, Stable Difusion, Azure CustomVison

## 5.자체 평가 의견

자체 평가

# 1.프로젝트 개요

## CONTENTS

# 근거 기반 설계

## 아동 학습 이론에 기반한 솔루션

글:림은 과학적 학습 이론에 기반한 서비스입니다. 특히 언어 발달 장애 아동의 특성을 고려해 인지심리학의 핵심 이론들을 적용했습니다. 글:림의 시각화 기능과 상호작용 유도는 이러한 이론적 토대에서 출발합니다.

글:림의 설계는 아동의 인지 발달과 학습 효율을 극대화하기 위한 연구 결과를 반영하여, 텍스트 이해의 장벽을 낮추고 즐거운 독서 경험을 제공합니다.

### 이중 부호화 이론 (Dual Coding Theory, Paivio)

인간은 언어(Verbal)와 시각(Visual) 정보를 각각 독립된 채널로 처리합니다. 앨런 파이비오의 연구에 따르면, 텍스트와 이미지가 결합될 때(이중 부호화) 정보 처리 효율과 기억력이 획기적으로 향상됩니다.

글:림 적용: 텍스트만 제시되는 환경은 아이들이 '시각적 채널'을 활용하지 못하게 하여 학습 효율을 떨어뜨립니다. 우리 서비스는 이 두 가지 채널을 동시에 자극하여 학습 효과를 극대화합니다.

### 인지 부하 이론 (Cognitive Load Theory, Sweller)

존 스웰러의 이론에 따르면, 학습자의 작업 기억 용량에는 한계가 있습니다. 추상적인 텍스트를 머릿속에서 이미지로 변환하는 과정은 불필요한 외재적 인지 부하(Extraneous Cognitive Load)를 발생시킵니다.

글:림 적용: 시각적 보조 자료(AI 생성 이미지)를 제공함으로써 아이들이 텍스트 해석에 쏟을 에너지를 줄여주고(인지 부하 감소), 대신 내용 이해와 상상에 집중하게 돕습니다.

### 연구 근거

Visual and Verbal Semantic Productions in Children with ASD, DLD... (2019)

해당 연구는 자폐 스펙트럼(ASD) 및 언어 발달 장애(DLD) 아동이 시각적 단서(Visual Cue)가 주어졌을 때 언어적 정보 습득과 회상 능력이 일반 아동과 유사한 수준으로 향상됨을 입증했습니다.

글:림 적용: 이는 우리 서비스가 제공하는 '시각적 리딩 메이트'가 단순한 흥미 유발을 넘어, 실질적인 언어 능력 향상의 도구가 될 수 있음을 시사합니다.



타겟 사용자 PROFILE

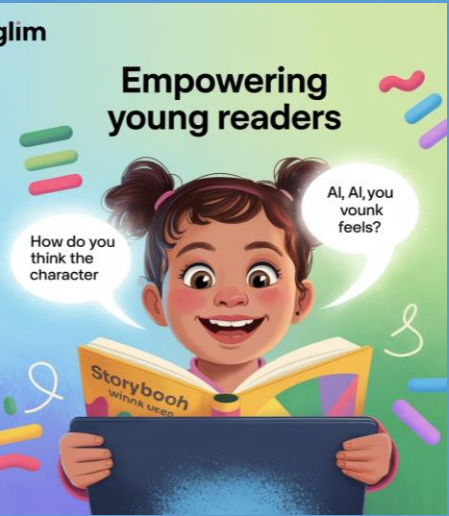
나이: 6-9세      특성: 언어 발달 지연 또는 장애      보조자: 부모/보호자      교육환경: 가정, 특수학급, 치료센터

아동 프로필

언어 발달 장애(ASD/DLD 포함) 가능성이 있는 아동으로, 텍스트 정보만으로는 상황을 구체화하거나 이미지를 연상하는 데 어려움을 겪습니다. 시각적 단서에 대한 반응이 우수하며, 반복 학습과 놀이형 교육에 긍정적으로 반응합니다.

부모/보호자 프로필

자녀의 언어 발달을 지원하기 위해 적극적으로 개입하고자 하는 부모/보호자로, 일상 속에서 효과적으로 활용할 수 있는 교육 도구를 찾고 있습니다. 디지털 기기에 대한 기본적인 이해를 갖추고 있으며, 자녀와의 상호작용 시간을 의미 있게 활용하고자 합니다.



아동의 니즈

텍스트를 쉽게 이해할 수 있는 시각적 변환, 자신의 감정과 생각을 안전하게 표현할 수 있는 환경, 흥미를 유지할 수 있는 반복 가능한 놀이형 학습 경험을 필요로 합니다.

부모의 니즈

짧은 시간 내에 높은 수준의 상호작용과 정서적 교류를 유도할 수 있는 도구, 자녀의 발달 상황을 객관적으로 관찰할 수 있는 기회, 전문가의 개입 없이도 가정에서 쉽게 활용할 수 있는 솔루션을 원합니다.

선호하는 학습 방식

시각적 도구와 이미지 중심 학습, 상호작용적 게임 활동, 즉각적인 피드백과 긍정적 강화, 관심 기반의 맞춤형 콘텐츠를 통한 학습에 더 집중합니다.

솔루션 기대 효과

자녀의 언어 이해력과 표현력 향상, 자발적인 상호작용 증가, 독서에 대한 흥미와 집중력 향상, 부모-자녀 간 질적인 소통 시간 확보, 장기적인 발달 지원이 가능한 지속 가능한 학습 경험을 기대합니다.

## 1.프로젝트 개요

# 시장 규모 언어장애·에듀테크

### 글로벌 언어 장애 시장

2024년 86.5억 달러 → 2031년 129.6억 달러 연평균 성장률(CAGR): 6.0%

자료출처: The Insight Partners(2024)

### 국내 에듀테크 시장

2021년 7.3조원 → 2023년 8.5조원 → 2024년 9.9조원 연평균 성장률(CAGR): 8.5%

자료출처: 산업연구원(2024)

### 한국 아동 재활 시장

언어·인지 재활 분야는 2024년 약 2억 달러 규모로 추정 2025-2035년 연평균 평균 성장률(CAGR): 6.66%

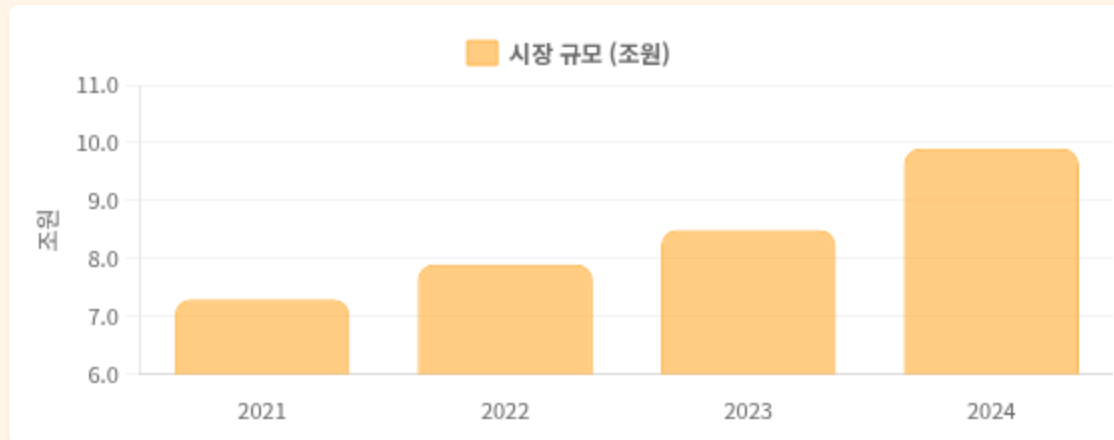
자료출처: Spherical Insights(2024)

글:림(GL:IM)

## 글로벌 언어 장애 시장 성장 전망



## 국내 에듀테크 시장 성장 추이



글:림의 시장 기회 언어 발달 장애 아동 대상 디지털 솔루션 시장이 지속 성장하는 가운데, 텍스트-이미지 변환 기술은 학습 보조 도구로서 높은 시장성을 보유하고 있습니다.

AI 기반 독서 리딩 메이트

# 경쟁사 분석

## 아이들나라 vs 글:림

시장 내 주요 경쟁 서비스 '아이들나라'와 '글:림'의 핵심 특징을 비교분석하여 차별화 포인트를 도출하고, 시장 진입 전략을 수립했습니다.

글:림은 '실시간 시각화'와 '무한 상상력'이라는 핵심 강점을 토대로 시장에서의 차별화된 포지셔닝을 확보할 수 있습니다.

구분	아이들나라 (LGU+)	글:림
서비스 모델	IPTV 부가서비스에서 OTT로 전환 올인원 키즈 콘텐츠 플랫폼	웹/앱 기반 AI 시각화 서비스 텍스트-이미지 변환 특화 솔루션
가격 정책	월 2만 5천원 정액제 U+TV 가입자 대상	월 9,900원~19,900원 구독제 Basic/Plus/Premium 티어
핵심 기술	AI 기반 추천 시스템 (ixi) TTS(음성 합성) 기술 활용	AI 비전 및 생성형 AI 실시간 텍스트→시각화
콘텐츠 접근	대규모 독점 콘텐츠 확보 (도서 전집, YBM, 디즈니 등)	어떤 텍스트에도 실시간 적용 무제한 확장성
차별화 포인트	브랜드 인지도 대규모 콘텐츠 라이브러리 AI 맞춤 추천 기능	N회차 독서 가능 다양한 시각적 해석 제공 Custom Vision 게임

### ⚠ 리스크 & 대응 전략

- IP·콘텐츠 협업 필요: 유명 IP와의 전략적 제휴를 통해 기존 캐릭터 활용 및 콘텐츠 확보
- 초기 친숙도 확보: 무료 체험판, 교육기관 파일럿 운영으로 사용자 확보 및 신뢰도 구축

# 유사 서비스 비교

## 국내외 사례 분석

현재 아동 독서 시장에는 다양한 앱과 서비스가 제공되고 있지만, 대부분은 정해진 콘텐츠를 소비하는 방식입니다. 글:림은 어떤 텍스트든 실시간으로 시각화하고, 대화와 게임으로 확장하는 새로운 접근법을 제시합니다.

같은 책을 다른 시각으로 계속 발견할 수 있는 것이 글:림만의 독보적인 가치입니다.

### 에픽(Epic)

미국

"아이들을 위한 넷플릭스"로 불리는 구독형 아동 도서 플랫폼으로, 4만 권 이상의 전자책 콘텐츠를 제공합니다. 독서 시간 추적, 배지, 퀴즈 등 게임화 요소가 특징이지만, 이미 제작된 고정 콘텐츠만 이용 가능합니다.

### 리드얼롱(Read Along)

구글

구글이 개발한 무료 읽기 연습 앱으로, AI 친구 '딩고'와 함께 책을 읽습니다. 발음 및 읽기 능력 개선에 초점을 맞추고 있으나, 시각적 상상력 확장 및 대화 유도 기능은 제한적입니다.

### 북스토리(Booxtory)

웅진씽크빅

AI 기반 독서 플랫폼으로, 독서 습관 형성과 사고력 향상을 유도하는 체험형 활동 서비스를 제공합니다. 그러나 특정 제휴 도서에 한정되어 있고, 독자의 자유로운 콘텐츠 확장이 어렵습니다.

### 글:림(GL:IM)

글:림

차별화 포인트: 어떤 텍스트든 실시간 맞춤형 시각화 + 자발적 대화 유도 질문 + 객체 인식 게임 기능이 결합된 유일한 서비스입니다. 특히 매번 다른 이미지가 생성되어 N회차 독서 경험이 가능합니다.

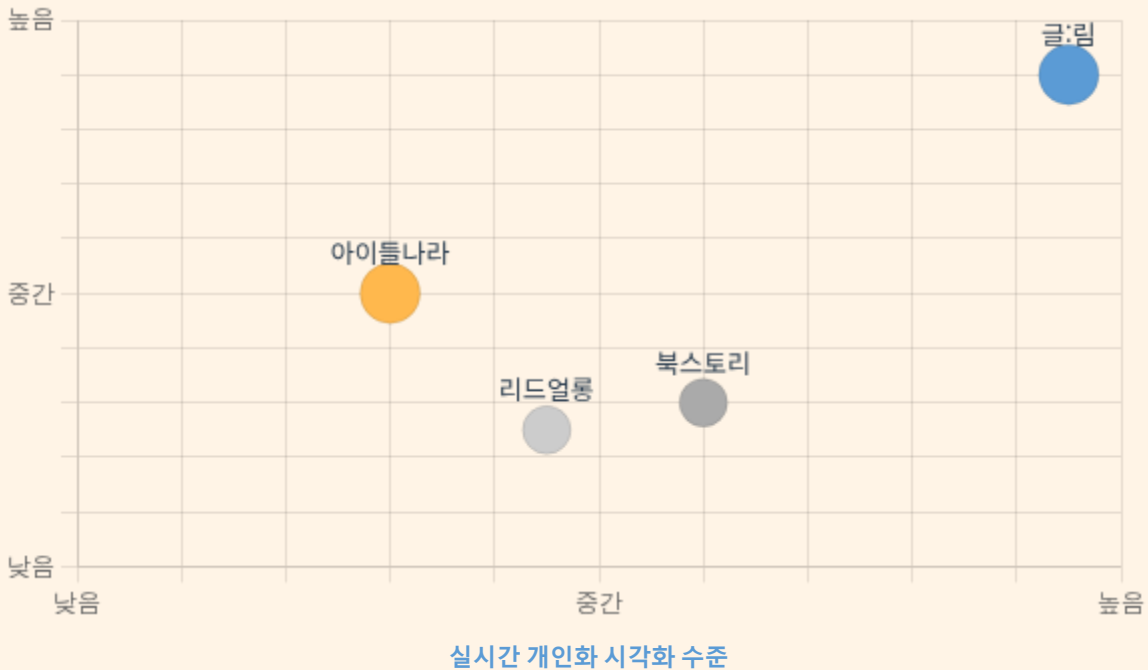
# 시장 포지셔닝 맵

글:림은 높은 수준의 실시간 개인화 시각화와 학습 상호작용을 결합하여, 기존 서비스들과 차별화된 위치를 점유하고 있습니다. 각 축이 나타내는 핵심 경쟁력을 기준으로 글:림과 다른 서비스들의 포지셔닝을 살펴볼 것입니다.

## 주요 포지셔닝 요소:

- x축: 실시간 개인화 시각화  
모든 텍스트를 즉각적으로 시각화하는 능력
- y축: 학습 상호작용 수준  
질문 생성, 게임화 등 능동적 참여 유도

## 독서 보조 솔루션 포지셔닝 비교



- 글:림  
실시간 텍스트→이미지 변환, 맞춤형 질문 생성, 게임 통합
- 아이들나라  
기존 콘텐츠 기반, AI 추천, 중간 수준 상호작용
- 리드얼롱  
낮은 개인화, 음성 인식 기반 상호작용
- 북스토리  
중간 수준 개인화, 제한된 상호작용

## 1.프로젝트 개요



# 무한 상상력

*One Story, Infinite Worlds*

핵심은 유지, 표현은 무한

같은 이야기라도 핵심 요소는 유지하면서 AI가 매번 다른 관점으로 시각화하여 보여줍니다. 아이가 읽을 때마다 새로운 세계를 발견합니다.

고정관념에서 자유로운 상상  
력

공주는 항상 금발? 마법사는 항상 노인? 전통적인 이미지에서 벗어나 다양한 표현과 시각을 경험하며 창의성이 자랍니다.

N회차 독서 동기 강화

같은 책도 매번 새롭게 읽는 즐거움을 통해 반복 독서를 자발적으로 유도하고, 이야기에 대한 더 깊은 이해와 애착을 형성합니다.

# 2.프로젝트 팀 구성 및 역할

## CONTENTS



## 2. 프로젝트 팀 구성 및 역할

Project Manger

박훈용 (일론-훈용)

- 프로젝트 비전 및 전략 수립
- 시장 조사 및 비즈니스 모델 설계
- 결과보고서 작성

Frontend

박지성 (사티아-지성)

- Web 클라이언트 개발
- UI 구현
- 시연 영상 제작

Product Manager

강민지 (머스크-민지)

- 서비스 아키텍처 설계 및 구축
- 백엔드 설계 및 개발

AI Engineer

안형주 (워렌-형주)

- Custom Vision 데이터 생성
- Stable Diffusion 최적화 설계
- 이미지 생성 로직 개발

AI Engineer

김성태 (나델라-성태)

- Custom Vision 모델 학습 및 최적화
- 데이터 파이프라인 구축
- 홍보 영상 제작

AI Engineer

송탁규 (버핏-탁규)

- Custom Vision 모델 학습 및 최적화
- 프롬프트 엔지니어링
- 데이터셋 생성 및 정제



# 개발 프로세스 & 10일간의 여정

## 사전 기획

11/12(수) ~ 13(목)

- 프로젝트 기획 및 주제 선정
- 팀 역할 분담
- 언어발달 장애아동을 위한 시각화 도구 컨셉 확정
- Azure 서비스 조합 아키텍처 설계
- 학술적 근거 수집

## 데이터 수집/전처리

11/13(금) ~ 17(화)

- Custom Vision 학습용 데이터 수집
- 데이터 라벨링 및 정제
- 이미지 편향성 발견, 프롬프트 규칙 확립
- 부적합 이미지 제거

## 모델 학습/최적화

11/17(월) ~ 19(수)

- Custom Vision 모델 학습 및 최적화
- SDXL 파라미터 튜닝
- Gemini API 프롬프트 엔지니어링
- 모델 성능 향상

## 서비스 구축 & 테스트

11/17(월) ~ 21(금)

- 서버 구축
- 백/프론트엔드 개발
- 통합 테스트 완료 및 시연 영상 제작

# 3.프로젝트 수행 절차 및 방법

## CONTENTS

# 기술 스택

## TECHNOLOGY STACK

글:림 서비스 구현을 위해 활용된  
클라우드, AI, 개발 및 협업 도구



### Cloud

Azure Vision(OCR): 텍스트 추출

Azure ML: 서버 구축

Azure Custom Vision: 객체 탐지

Google Gemini: 프롬프트 생성



### AI 모델

Stable Diffusion XL 1.0: 이미지 생성

Stable Diffusion LoRA: 이미지 생성(테스트)



### 백엔드 & 프론트엔드

Python: 백엔드 개발

JavaScript: 프론트엔드

FastAPI: API 서버

Ubuntu VM: 서버 환경



### 협업 & 디자인

Canva: 웹페이지 그래픽 디자인

Vrew: 시연 영상 편집

Design.com: 로고 및 UI 디자인

Microsoft Teams: 팀 협업

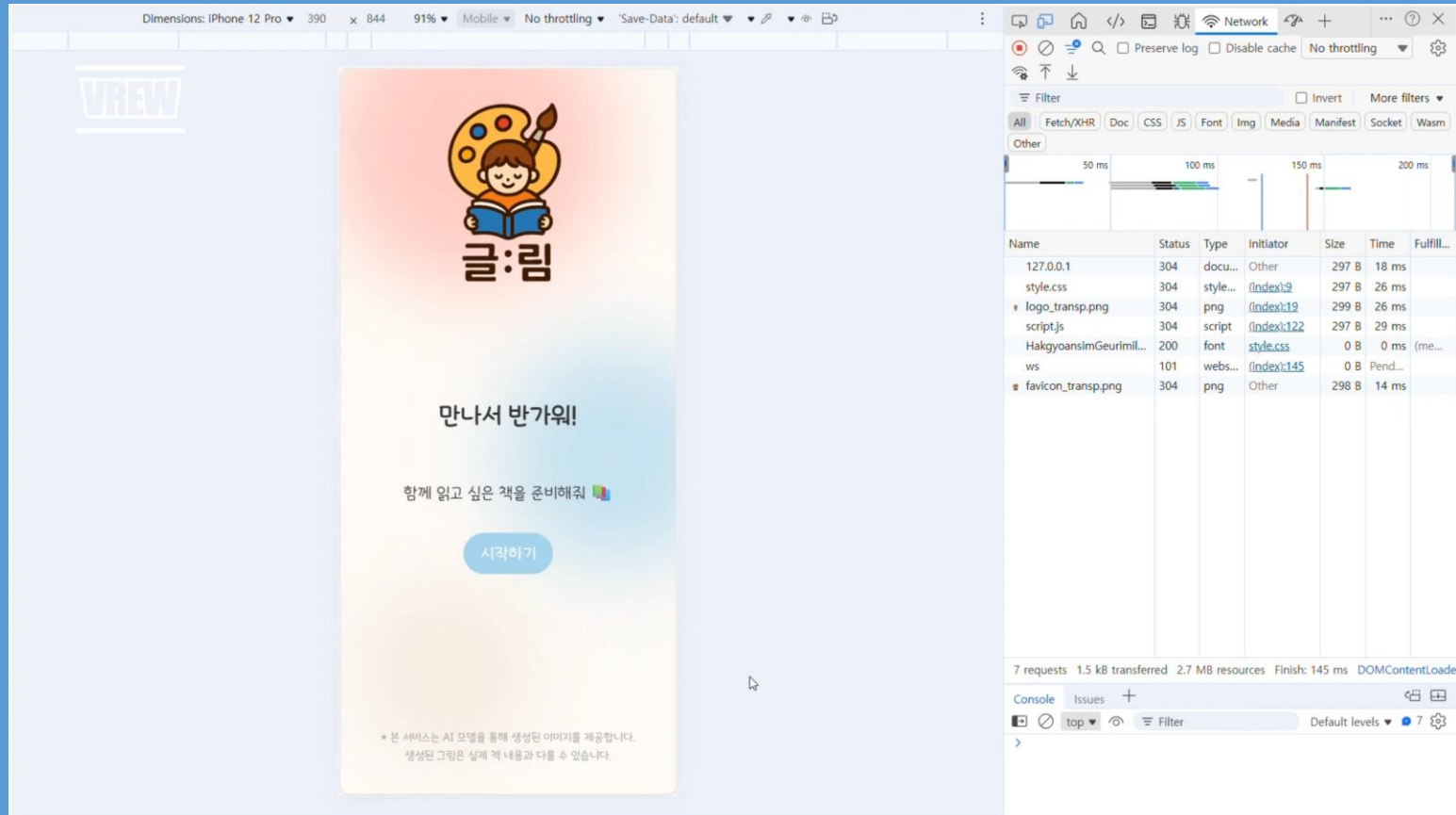
# 시스템 아키텍처



# 4.프로젝트 수행 결과

## CONTENTS

# 서비스 . UI/UX



# 서비스 기능

3단계 프로세스로 아이의 독서 경험을 풍부하게

## 1 입력


책표지나 본문을 촬영하면 Azure OCR이 자동으로 텍스트를 추출합니다. 아이가 읽고 있는 동화책이나 교과서의 내용을 쉽고 빠르게 인식합니다.

## 2 시각화 · 대화

Gemini API가 추출된 텍스트를 분석하여 SDXL을 통해 생생한 이미지를 생성합니다. 동시에 내용 이해와 감정 표현을 돕는 5가지 질문을 제시하여 자연스러운 대화를 유도합니다.


## 3 놀이

Azure Custom Vision의 Object Detection 기술로 생성된 이미지 속 사물을 인식해 그림찾기 게임을 자동으로 생성합니다. 재미있는 상호작용을 통해 아이의 주의 집중력과 어휘력을 향상시킵니다.




글:림

만나서 반가워!

함께 읽고 싶은 책을 준비해줘 

시작하기

\* 본 서비스는 AI 모델을 통해 생성된 이미지를 제공합니다.  
생성된 그림은 실제 책 내용과 다를 수 있습니다.




글:림

어떤 책을 읽을까?

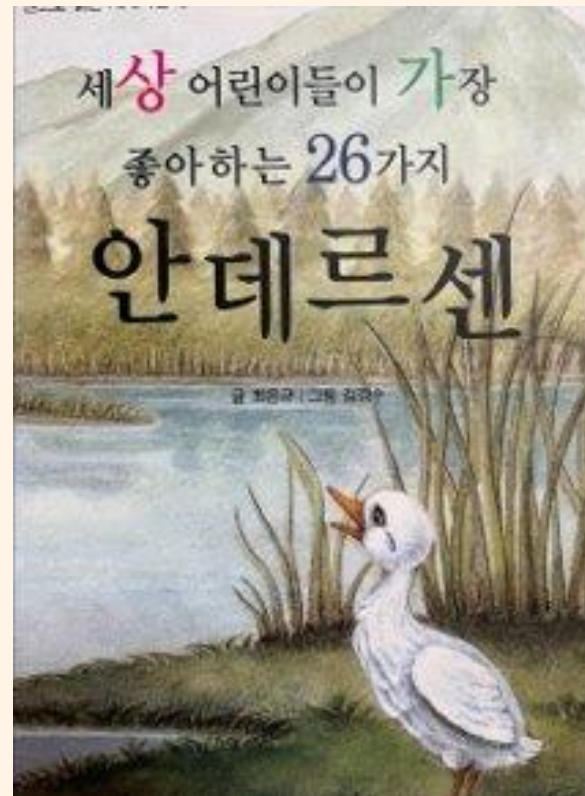
책 제목

여기에 제목을 입력해줘

또는

 책 표지 촬영/업로드

다음




글:림

어떤 책을 읽을까?

책 제목

세상 어린이들이 가장 좋아하는 26가지 안데르센

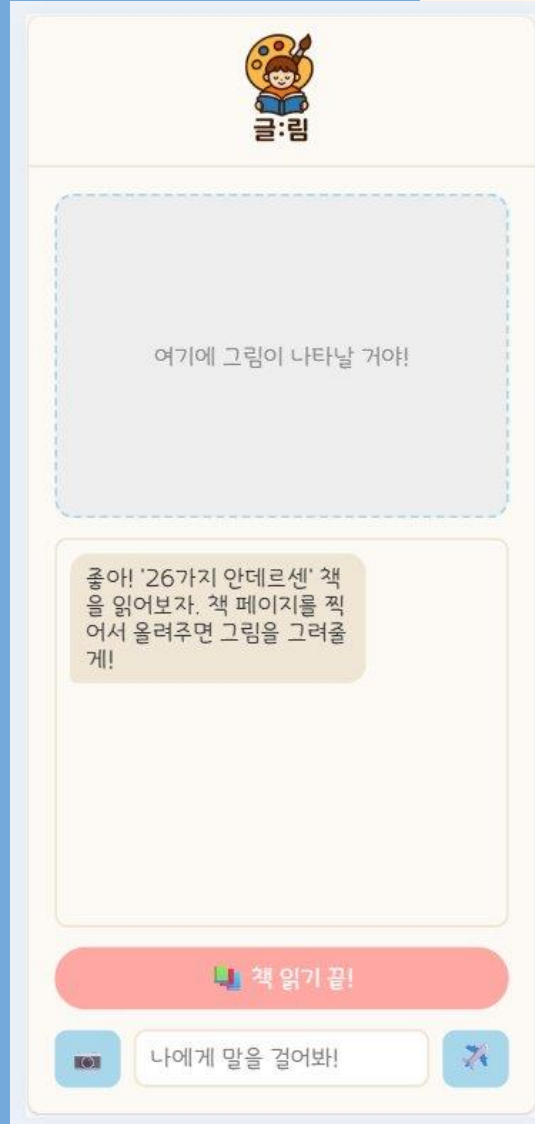
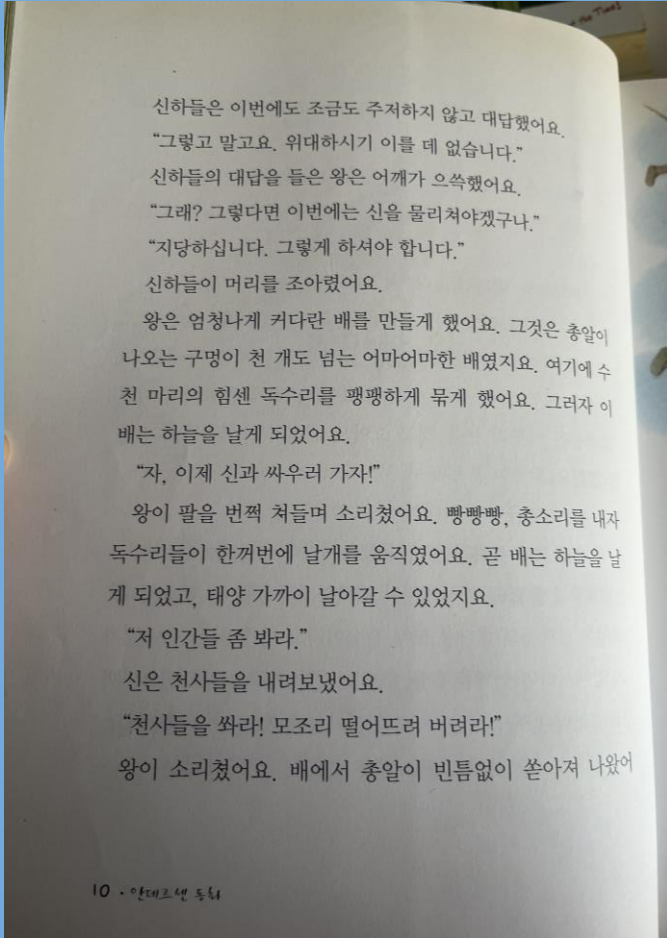
또는

 책 표지 촬영/업로드

다음



# 1 입력



ocr\_text: 신하들은 이번에도 조금도 주저하지 않고 대답했어요.  
 "그리고 말고요. 위대하시기 이를 데 없습니다."  
 신하들의 대답을 들은 왕은 어깨가 으쓱했어요.  
 "그래? 그렇다면 이번에는 신을 물리쳐야겠구나."  
 "지당하십니다. 그렇게 하셔야 합니다."  
 신하들이 머리를 조아렸어요.  
 왕은 엄청나게 커다란 배를 만들게 했어요. 그것은 총알이 나오는 구멍이 천 개도 넘는 어마어마한 배였지요. 여기에 수 천 마리의 힘센 독수리를 팽팽하게 묶게 했어요. 그러자 이 배는 하늘을 날게 되었어요.  
 "자, 이제 신과 싸우러 가자!"  
 왕이 팔을 번쩍 쳐들며 소리쳤어요. 뽕뽕뽕, 총소리를 내자 독수리들이 한꺼번에 날개를 움직였어요. 곧 배는 하늘을 날 게 되었고, 태양 가까이 날아갈 수 있었지요.  
 "저 인간들 좀 봐라."  
 신은 천사들을 내려보냈어요.  
 "천사들을 쏘라! 모조리 떨어뜨려 버려라!"  
 왕이 소리쳤어요. 배에서 총알이 빈틈없이 쏟아져 나왔어

# 1 입력

OCR API 관련 시행착오 → 알고 보니 "이미지 사이즈 문제"

클라우드/API 서비스를 이용할 때 **Requirements**를 잘 읽어 봐야 한다는 것을 다시금 깨달았습니다.

## Service limits

### Input requirements

Version 4.0

Version 3.2

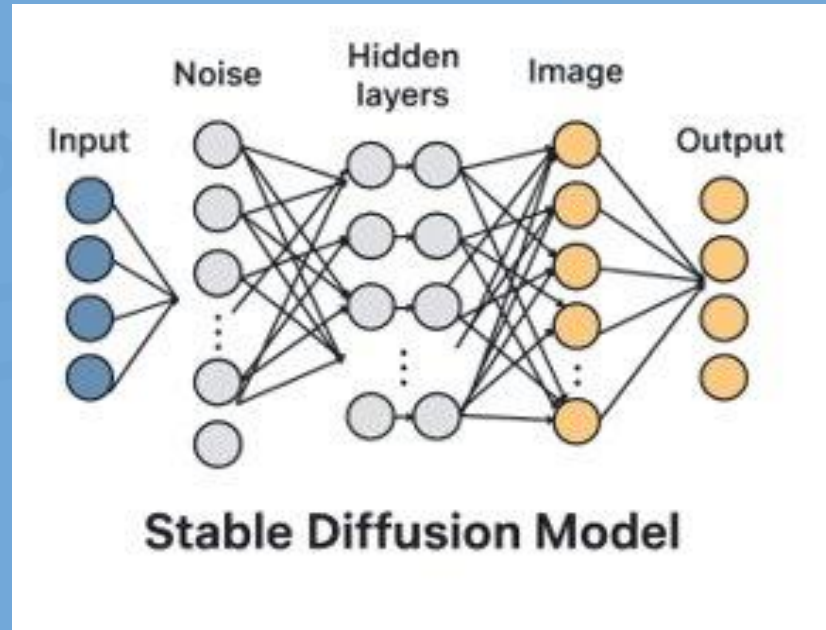
Image Analysis works on images that meet the following requirements:

- The image must be in JPEG, PNG, GIF, or BMP format
- The file size of the image must be less than 4 megabytes (MB)
- The dimensions of the image must be greater than 50 x 50 pixels and less than 16,000 x 16,000 pixels

## 2 ✎ 시각화 · 대화

### 스테이블 디퓨전 이미지 생성 흐름

프롬프트



## 2 ✏️ 시각화 · 대화

### 아이와의 상호작용

#### - 제미나이 연동



  
글:림





왕이 만든 배는 "어마어마한 배"라고 했는데, 네 생각에 그 배는 얼마나 컸을까? 무슨 색이었을 것 같아?

 책 읽기 끝!





  
글:림



황금색!

응, 황금색이구나! 반짝반짝 빛나서 정말 예쁘겠다. 또 어떤 색깔이 보이는 것 같아?

구름색도 보이고, 음... 넌 뭐가 보여?

와, 구름색이라니! 정말 멋진 표현이다. 선생님은 뭔가 반짝이는 게 보이는 것 같기도 한데, 너는 또 뭐가 보여?

 책 읽기 끝!







## 2 ✏️ 시각화 · 대화

### 아이의 시각적 이해 촉진을 유도하는 프롬프트 설계

question\_prompt = """ **\*\*Role:\*\*** You are a highly specialized AI designed to serve as the core function of an **\*\*'AI-Powered Reading Companion'\*\*** for young children. **\*\*Objective:\*\*** Your primary task is to analyze a given children's story text (fairytale/picture book content) and generate data that specifically aids the language comprehension and developmental needs of children aged **\*\*3 to 7 years old\*\***, who require visual information for better understanding. **\*\*Core Output Directives:\*\*** You must produce a structured output focused on promoting active engagement and linguistic development. --- **### \*\*Language Development Questions (5 Total)\*\*** \* Generate **\*\*exactly five (5) high-quality questions\*\*** to facilitate an engaging dialogue with the child.

```
ai_question: Q1. 왕은 하늘을 나는 배를 타고 누구랑 싸우러 간다고 했지?  
Q2. 신하들이 왕에게 "위대하시기 이를 데 없습니다!" 하고 말하니까 왕은 기분이 어땠을까? 왜 그렇게 생각했어?  
Q3. 왕이 천사들을 쏘라고 했는데, 그 다음에는 어떤 일이 일어날 것 같아?  
Q4. 왕이 만든 배는 "어마어마한 배"라고 했는데, 네 생각에 그 배는 얼마나 컸을까? 무슨 색이었을 것 같아?  
Q5. 만약 네가 왕이라면, 그 멋진 하늘을 나는 배를 타고 뭘 하고 싶어?
```

## 2 ✏️ 시각화 · 대화

### 아동 안전성 강화, 스테이블 디퓨전의 문제점 개선 프롬프트 설계

system\_prompt = "" Role: You are an expert prompt engineer for Stable Diffusion who turns children's story text into consistent, detailed illustration prompts. Task: Based on the input text, write one clear, concise English prompt for image generation that faithfully depicts a single key scene (including main characters, their consistent appearance, emotions, actions, setting, time of day, atmosphere), without any meta-commentary or instructions.

CRITICAL RULES (Must Follow):

1. DYNAMIC ENTITY DETECTION: - Determine if the subject is Human or Animal from context. - BIAS FIX: If the subject is a human role (e.g., King, Queen, Student), MUST add "Human" prefix (e.g., "Human King"). - ANIMAL FIX: If it is an animal, specify the species clearly (e.g., "Baby Bear animal").
2. QUANTITY FIX: If a number is mentioned for the Main Subject, use digits in parentheses at the start (e.g., "(5) baby ducks").
3. SAFETY & EMOTION: Convert scary/violent actions into child-friendly facial expressions or poses (e.g., "crying" -> "sad face", "fighting" -> "standing confidently").

negative\_prompt = ("deformed face, distorted face, asymmetrical face, extra eyes, extra limbs, " "extra fingers, long neck, disfigured, mutated, low quality, blurry, distorted, " "text, cropped, ugly, disfigured, poor anatomy, missing limbs, malformed hands, " "poorly drawn eyes, unsettling, monochrome, grayscale, realistic, photography, photo")

```
sd_prompt: Mighty, proud Human King on a colossal flying ship propelled by (thousands of)
powerful eagles animal, standing confidently as the ship soars towards the sun, bright cl
ear sky with (several) descending angels approaching.
```

## 2 ✏ 시각화 · 대화

### 1. Stable Diffusion 환경 구축

#### (1) SDXL 환경 설정

- Base / V1.5 / Refiner / LoRA 설치

#### (2) 이미지 품질 개선

- 모델별 비교(SDXL v1.0/1.5, Refiner, LoRA)
- 설정값 및 품질 실험

The screenshot displays the 'Generation' tab of the Stable Diffusion web interface. The settings are as follows:

- Sampling method:** DPM++ 2M
- Schedule type:** Karras
- Sampling steps:** 30
- Hires. fix:** ☐
- Refiner:** ☒ Refiner
- Checkpoint:** sdxl\_base.safetensors [31e35c80fc]
- Switch at:** 0.8
- Width:** 1024
- Height:** 1024
- Batch count:** 1
- Batch size:** 1
- CFG Scale:** 10
- Seed:** -1
- Extra:** ☐

## 2 ✎ 시각화 · 대화

### 2.데이터셋 구축

#### (1) 이미지 데이터셋 생성

- SDXL로 500장 이상 생성
- 설정값 실험 및 프롬프트 품질 비교에 활용





## 2 ✏️ 시각화 · 대화

### 3.프롬프트 엔지니어링

#### (1) Gemini 시스템 프롬프트 설계

#### (2) 문장 → 프롬프트 변환 규칙

- 장면 중심화
- 핵심 요소 자동 포함
- 설명 없는 순수 프롬프트 출력

#### Human / Animal 구분 규칙

King → Human King

Bear → Bear animal

#### 수량 규칙

(3) ducks

#### 감정/행동 안전 변환 규칙

fighting → standing confidently

#### 스타일 템플릿 유지

동화풍 스타일을 일관되게 유지하며, 설명문이 아닌 순수 프롬프트만 출력합니다.  
Negative Prompt를 통해 기괴한 요소, 기형, 현실 사진처럼 보이는 요소를 제거합니다.

Positive: whimsical, children's book illustration style

Negative: deformed face, extra limbs, text, blurry, distorted

# Custom Vision 학습과정

## 데이터 셋 구성과 훈련 최적화

### 초기 학습 단계

Q1

Custom Vision의 Tagging 과정은 원활하게 진행

Stable Diffusion으로 생성한 이미지만 사용

단일 데이터 소스로 인한 과적합(overfitting) 우려 발생

### 데이터 셋 확장

Q2

Ot & Sien 데이터셋으로 확대

다양한 데이터 소스 추가로 인한 혼란 증가

데이터 통합 과정에서 일관성 문제 발생

### Iteration 증가 시도

Q3

Iteration을 15까지 점진적으로 증가

Precision, Recall, mAP 지표 수치 여전히 낮음

단순 반복 학습으로는 성능 개선 불가 판단

### 훈련 방식 변경

Q4

Quick Train → Advanced Train으로 전환

훈련 시간 약 3시간으로 대폭 증가

시간 투자에도 불구하고 평가지표 개선 미미

### 문제 원인 분석

Q5

훈련 데이터셋 이미지가 과도하게 오래된 느낌

Stable Diffusion 생성 이미지와 스타일 부합

데이터 품질 불균형으로 성능 저하의 주요 원인

# 프로젝트 수행 경과

- 동화원문 단락 별로 Gemini에게 질의 응답 형식 생성 (JSON)
- 동화책 3권 (거만한 왕, 기름장수네 난장이, 미운 아기 오리)
  - questions : 3가지 (내용 이해, 감정 추론, 맞춤형대화유도)
  - image\_keywords : 인상적인 장면 /구체적•시각적 이미지
  - game\_object : 학습 보조 게임용 핵심 명사 추출
- 동화원문에서 stable-diffusion 프롬프트용 키워드 추출 후 이미지 생성  
784개(키워드 49 set X 1set당 16개)
- Ot & Sien dataset( 네덜란드 유명 동화책의 이미지 dataset )
- 1,512 images with 7,210 objects 에서 786개 엄선

## JSON 출력물

```
{
  "questions": [
    "엄마 오리가 '야호!'하고 소리쳤을 때 기분이 어땠을까요? 엄마 오리의 표정이나 마음을 이야기해 볼까요?",
    "새끼 오리들이 '꽹꽹'하고 소리를 냈는데, 이 소리는 오리가 언제 내는 소리일까요? 다른 동물들은 어떻게 소리를 내는지 말해줄래요?",
    "만약 여러분이 아기 오리들을 처음 본다면, 어떤 말을 해주고 싶나요?"
  ],
  "image_keywords": [
    {
      "scene_description": "엄마 오리가 앞에서 깨어나는 새끼들을 보는 장면",
      "keywords": [
        "따뜻한 햇살이 비치는 둥지",
        "행복해 보이는 엄마 오리의 미소",
        "알껍데기에 금이 간 모습",
        "갓 태어난 노란 새끼 오리들"
      ]
    }
  ],
  "game_objects": [
    "아기오리"
  ]
}
```

# 프로젝트 수행 경과 1

## 1. Python SDK 기반 데이터 파이프라인 구축

- Azure Custom Vision SDK와 Python을 연동하여 대량의 이미지 데이터 업로드 자동화
- 수작업 대비 데이터셋 구성 시간 단축 및 효율성 확보

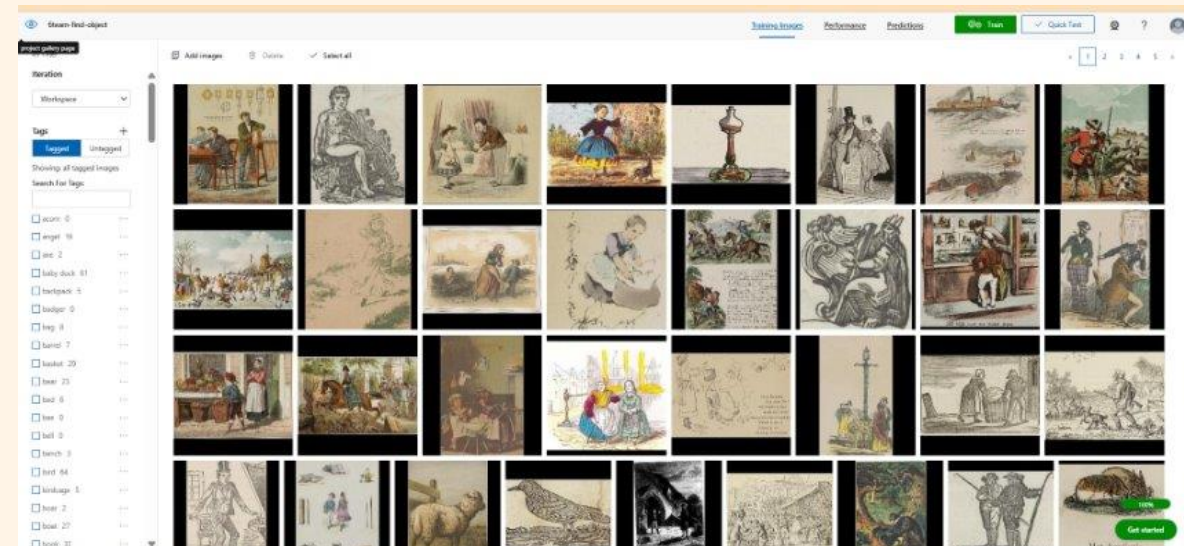
### 1.1 초기 문제점 및 원인 분석:

**문제점:** 초기 테스트에서 트라이얼 횟수(Iteration)를 거듭했음에도 불구하고, 전반적인 성능(Recall/Precision)이 정체되거나 불안정한 현상이 발생

### 원인 분석:

- **성능 정체:** 초기 트레이닝 반복(Iteration)에도 Precision/Recall 성능이 불안정하게 정체되는 현상 발생.
- **데이터 질 문제:** 학습 데이터셋 내에 저품질 이미지(Low Quality)가 포함되어 모델의 학습을 방해함.
- **편향성 학습:** 재해(재난) 상황 데이터의 수량 불균형(Data Imbalance)으로 인해, 모델이 특정 태그에만 치우쳐 학습되는 편향성 발생.

```
1 # Azure Custom Vision SDK와 Python을 연동하여 이미지 데이터를 일괄 업로드하는 스크립트입니다.
2 import os
3 import time
4 from azure.cognitiveservices.vision.customvision.training import CustomVisionTrainingClient
5 from azure.cognitiveservices.vision.customvision.training.models import ImageFileCreateBatch
6 from msrest.authentication import ApiKeyCredentials
7
8 # ===== 설정값 입력 =====
9
10 # 1. Azure Custom Vision 인스턴스 정보
11 TRAINING_KEY = "YOUR_CUSTOM_VISION_TRAINING_KEY"
12 ENDPOINT = "YOUR_CUSTOM_VISION_ENDPOINT_URL"
13
14 # 2. 프로젝트 및 데이터 정보
15 PROJECT_ID = "YOUR_CUSTOM_VISION_PROJECT_ID"
16 # 업로드할 이미지가 저장된 로컬 경로
17 # 예시: 'data/' 폴더 안에 'tag_A/'와 'tag_B/' 폴더가 있다면, 'tag_A'와 'tag_B'가 태그 이름이 됩니다.
18 IMAGE_FOLDER_PATH = "./images_to_upload"
19
20 # ===== 실행 =====
21
22 def upload_images_from_folder(training_client, project_id, image_folder_path):
23     """
24     로컬 폴더를 스캔하여 폴더 이름을 태그로 사용하고 이미지를 일괄 업로드합니다.
25     """
26     print(f"--- 프로젝트 ID: {project_id} 에 이미지 업로드를 시작합니다. ---")
27
28     # 1. 폴더 스캔 및 태그/이미지 매핑
29     all_image_files = {}
30
31     # 이미지 폴더 경로가 존재하지 않으면 종료
32     if not os.path.exists(image_folder_path):
33         print(f"[오류] 지정한 이미지 폴더 경로가 존재하지 않습니다: {image_folder_path}")
34         return
35
36     # IMAGE_FOLDER_PATH 내부의 모든 서브디렉토리 순회
37     for tag_name in os.listdir(image_folder_path):
38         tag_path = os.path.join(image_folder_path, tag_name)
```



# 프로젝트 수행 경과 2

## 2. 재현율(Recall) 향상 노력의 필요성 (Why do we focus on Recall?)

재현율(Recall)이 중요한 이유? False Negative 방지!

**재현율(Recall) 정의:** 실제 '정답'인 것들 중에서 모델이 얼마나 많이 '정답'이라고 맞췄는지 나타냄

**Recall 하락 시 영향:** 거짓 음성(False Negative, FN)이 발생할 경우, 텍스트에 분명히 있던 사물(예: 빨간 사과)을 모델이 인식하지 못해 AI가 생성한 이미지에서 누락

**서비스 가치 훼손:** 핵심 오브젝트 누락은 AI와의 대화/미니 게임 유도 실패로 이어져 서비스의 기본 목적과 흥미를 심각하게 떨어뜨림(가장 치명적인 오류)

초기 문제 해결을 위한 데이터 전략 (Recall 향상 방안)  
초기 테스트에서 발생했던 '저품질 이미지' 및 '데이터 불균형' 문제를 해결하고 재현율을 높이기 위해 다음 3가지 데이터 개선 작업을 진행.



전략 구분	초기 문제점	개선 조치 및 결과
데이터 일관성 확보	기존 데이터셋 내 오래된 사진 및 저품질 이미지(Low Quality) 포함으로 모델 학습 방해.	<b>Stable Diffusion 이미지로 통일:</b> 데이터셋에서 오래되거나 저품질의 이미지를 모두 제거하고, <b>Stable Diffusion</b> 으로 생성된 고화질 이미지만 사용했
데이터 다양성 확보	항목별 데이터 수량의 불균형(Data Imbalance)으로 모델이 소수 항목을 놓치는 편향성 발생.	<b>항목별 수량 일치:</b> 각 태그 항목(등장인물, 사물 등)의 <b>이미지 수량을 동일하게(Equalization)</b> 맞추어 편향성을 제거하고 모든 항목에 대해 균등한 학습을 유도
Recall 결과	(Iteration 15 결과)	<b>Recall 95.1% 달성:</b> 치명적인 누락(FN)을 최소화하여 서비스의 기본 가치를 확보 <small>AI 기반 독서 리딩 메이트</small>



# 프로젝트 수행 경과 3

## 3. Main importance : '양'보다는 '질'과 '균형'

- 데이터 정제(Cleaning): 객체가 불분명하거나 노이즈가 심한 이미지 제거
- 데이터 밸런싱(Balancing): 각 카테고리(Class)별 학습 데이터 개수를 균일하게 맞추어 편향 제거
- 결과: 데이터 품질 개선 후, 모델 성능의 비약적 향상 확인

## 4. 성능 개선 결과(Iteration 10[초기 모델]vs 15[완성본])

- Recall (재현율): 65.5% → 95.1% (▲29.6%p)
- mAP (평균 정밀도): 66.6% → 98.2% (▲27.5%p)
- Precision(예측률) 43.3%→85.8%(▲42.5%p)
- 데이터 최적화를 통해 실서비스 가능한 수준의 모델 성능 확보

### Iteration 10

Finished training on 2025. 11. 17. 오후 3:37:02 using General [A1] domain  
Iteration id: 544c270d-8cf4-4694-9b6e-5526616ce2ad



### Iteration 15

Finished training on 2025. 11. 19. 오후 4:34:43 using General [A1] domain  
Iteration id: d910d245-5621-4558-a157-f80603327dab  
Published as: Iteration15



# 5. 자체 평가 의견

## CONTENTS

# 자체 평가 의견

현재 아동 독서 시장에는 다양한 앱과 서비스가 제공되고 있지만, 대부분은 정해진 콘텐츠를 소비하는 방식입니다. 글:림은 어떤 텍스트든 실시간으로 시각화하고, 대화와 게임으로 확장하는 새로운 접근법을 제시합니다.

같은 책을 다른 시각으로 계속 발견할 수 있는 것이 글:림만의 독보적인 가치입니다.



## 사전 기획의 관점에서 프로젝트 결과물에 대한 완성도 평가(10점 만점)

8점

배포를 하지 못함  
구현하고자 하는 기능을 전부 구현하지 못함



## 개인 또는 우리 팀이 잘한 부분과 아쉬운 점

잘한 점: 역할분담이 잘 되었다고 생각한다. (기획/프론트엔드/백엔드/SD/CV)  
아쉬운 점: 다양한 애저 프레임워크로 실험해보고 싶었으나 시간 부족



## 프로젝트 결과물의 추후 개선점이나 보완할 점 등 내용 정리

Image to image 기능 추가  
음성인식 기능 추가



## 프로젝트를 수행하면서 느낀 점이나 경험한 성과(경력 계획 등과 연관)

서비스 기획을 하고 실제로 구현되는 것을 경험해보니, 신기하기도 아쉽기도 한 느낌.  
AI의 발전으로 인한 엄청난 생산성 향상이 피부로 와 닿았다.



5. 자체 평가 의견

항목	적용여부	진단	보안 및 강화 방안
투명성	O	생성형 AI 결과물임을 명확히 밝히지 않으면, 진짜 그림책 작가의 저작물로 오해할 수 있음.	서비스 첫 화면에 ***이 그림은 인공지능이 생성한 가상의 이미지입니다.***라는 <b>안내 문구 고지</b> .
책임성	X	AI가 생성한 이미지로 인해 사용자에게 심각한 문제(예: 부적절한 이미지 생성)가 발생했을 경우에 대한 보상 및 책임 원칙이 없음.	서비스 이용 약관에 **'의료 목적이 아닌 참고용으로만 사용해야 함'***을 명시하고, AI의 오작동으로 인한 면책 범위를 설정. 다만, 서비스 내부 결함으로 인한 피해 발생 시 <b>책임 및 보상 절차</b> 를 마련하여 고지.
공정성	O	아동용 서비스의 경우, 편향된 이미지가 고정 관념을 형성할 위험이 큼. (예: 왕은 항상 백인)	이미지 생성시 시스템 프롬프트 값에 편향 조절 장치를 마련 (결과: 왕은 백인, 흑인 교차 생성) 또한 '상상력버튼' 기능을 추가하여, 의도적으로 성별, 인종, 연령을 다양하게 교차 생성하여 편향된 결과가 고착되는 것을 방지.
신뢰성 및 안전성	O	Gemini, SDXL 등 외부 API 의존도가 높으며, 내부에서 생성하는 결과물에 대한 모니터링 체계가 미흡함.	<b>프롬프트 필터링 강화:</b> 사용자 입력 텍스트와 Gemini가 생성한 SDXL용 프롬프트 모두에 2단계 유해 콘텐츠 필터(Harmful Content Filter)를 적용하여 부적절한 이미지 생성 시도를 원천 차단.
포용성	O	현재는 한국만 서비스 출시 텍스트를 주로 처리, 서비스 대상 지역 확장 이미 고려 완료했음	<b>다국어 OCR 및 Prompt 처리 기능 확보(예정):</b> 현재 Gemini API 활용으로 다국어로 손쉽게 확장 가능
개인정보 보호 및 보안	해당안됨 (배포X)	입력된 이미지 데이터(책 사진)와 생성된 그림(자서전)은 민감한 개인 기록이므로 유출 위험에 대비해야 함.	<b>저장 데이터 암호화:</b> 사용자의 텍스트 입력과 생성된 이미지 데이터는 <b>암호화된 Firestore</b> 에 저장하며, <b>사용자 인증(UID)</b> 기반으로만 접근 가능하도록 설정.

# 감사합니다

글:림 프로젝트에 관심을 가져주셔서 감사합니다.  
언어 발달 장애 아동들에게 독서의 즐거움을 선사하고 부모와의 소통을 도울 수 있는 이 서비스가 많은 아이들에게 도움이  
되기를 바랍니다. 텍스트를 이미지로, 추상적인 글자를 선명한 세상으로 연결하는 여정에 함께해 주세요.

팀 : 육안(肉眼) 6조

이메일 : [glim@mail.com](mailto:glim@mail.com)

웹사이트 : [www.glim-reader.com](http://www.glim-reader.com)

소셜 미디어 : [@glim\\_reader](https://www.instagram.com/glim_reader)